Article

Natural Virtual Reality User Interface to Define Assembly Sequences for Digital Human Models

Andreas Geiger [1,*], Elisabeth Brandenburg [1] and Rainer Stark [1,2,*]

[1] Virtual Product Creation, Fraunhofer IPK, Pascalstraße 8-9, 10587 Berlin, Germany; elisabeth.Brandenburg@ipk.fraunhofer.de

[2] Industrial Information Technology, Technische Universität Berlin, Pascalstraße 8-9, 10587 Berlin, Germany

* Correspondence: andreas.geiger@ipk.fraunhofer.de (A.G.); Rainer.Stark@tu-berlin.de (R.S.)

check for updates

Abstract: Digital human models (DHMs) are virtual representations of human beings. They are used to conduct, among other things, ergonomic assessments in factory layout planning. DHM software tools are challenging in their use and thus require a high amount of training for engineers. In this paper, we present a virtual reality (VR) application that enables engineers to work with DHMs easily. Since VR systems with head-mounted displays (HMDs) are less expensive than CAVE systems, HMDs can be integrated more extensively into the product development process. Our application provides a reality-based interface and allows users to conduct an assembly task in VR and thus to manipulate the virtual scene with their real hands. These manipulations are used as input for the DHM to simulate, on that basis, human ergonomics. Therefore, we introduce a software and hardware architecture, the VATS (virtual action tracking system). This paper furthermore presents the results of a user study in which the VATS was compared to the existing WIMP (Windows, Icons, Menus and Pointer) interface. The results show that the VATS system enables users to conduct tasks in a significantly faster way.

Keywords: virtual reality; leap motion; DHM; user study; VATS; user interaction; grasping

1. Introduction

The goal of Model Based System Engineering (MBSE) is to consider all system-relevant parameters within the virtual product creation. Shortening the product development process and thus saving money is motivation for a holistic model mapping. Once the product to be developed and additionally dependent subsystems have been developed, later dependencies and interactions can be predicted. In particular, advanced systems engineering (ASE) emphasizes the protection of the user and postulates the protection of human factors in the early phases of product development.

The evaluation of human factors is not new. Aspects such as user experience and usability are, for example, required as sales arguments. Still, it is often the case that human factors are usually only tested on real prototypes in late phases of product development. In this phase, significant changes to the product are hardly possible, and it is therefore necessary to efficiently integrate the assurance of human factors into the early virtual product development process.

One of the central challenges is to involve human beings already in the early stages of the product development process to design suitable assembly stations and ensure consistent product quality [1]. Therefore, companies use digital human models (DHMs), which are digital representations of human beings. They are used to design secure and efficient workplaces and processes [2]. Such DHMs are often integrated into computer-aided software systems (JACK by Siemens PLM (Siemens PLM 2016), RAMSIS by Human Solutions [3] as well as the HUMAN from Dassault Systèmes [4]). The use of DHM software improves the communication possibilities within companies [5]. The reason why many companies do not use DHMs is the high cost of training engineers to use them for ergonomic analysis.

Thus, the complexity of the software keeps small and middle-sized enterprises away from using DHMs. The report "Digital ergonomics 2025" [6] describes the results of a survey with experts about trends of digital human models until 2025. The experts recognized six impediments of DHM software systems, of which three address the aspect of usability of such software systems: the high complexity of the software systems, the high amount of time to conduct virtual simulations, and the high operating costs, especially to train the employees. There are already DHM software systems that attempt to address these impediments and put a focus on optimized usability. For example, the digital human model "Intelligently Moving Manikin" (IMMA) [7] provides collision-free path and movement planning, which enables users to analyze assembly processes and conduct more precise and efficient ergonomic assessments. They increase the usability by decreasing configuration effort to use manikin families. A manikin family is a group of manikins "intended to accommodate the population based on a certain confidence level." [8]. Such manikin families include data about anthropometrics and the gender of the population of different regions of the world.

One way to reduce the complexity of using DHM software is to make use of the possibility of working with DHMs in a virtual environment intuitively [6]. Due to today's democratization of VR technology, it is possible to integrate VR technologies like HMDs into the daily workplace at a low cost [9]. VR offers the possibility of representing 3D models on a one-to-one scale [10]. This representation already integrates the human perspective into the product development process [11]. Users are thus able to experience the environment and investigate construction models in design reviews [12]. Software companies like MiddleVR [13] or IC.IDO [14] provide software tools, which allow the use of DHMs in a VR-application. However, this type of software only allows for conducting ergonomic assessments in two ways.

The first is using motion capture systems. This method allows the recording of body movements and mapping to a digital human model, which is then used to simulate the ergonomics. However, it is not possible to use the individually captured motion for different manikin percentiles. Thus, assembly planners can only assess the ergonomic criteria of the person who was performing the assembly task.

The second is to put the DHM in static positions to, for example, conduct reachability studies by pulling the manikin into the right position with controllers. It enables engineers to investigate the reachability study for whole manikin families. Among others, like Siemens Technomtix, and Dassault Systèmes Jack/Jill, IPS IMMA provides a motion capture and VR user interface.

The design of interaction dialogs with software or technical applications requires a user-centered and task-appropriate method. Only then can complex tasks be processed efficiently. Above all, the design of interaction in VR has to be application-specific. On the one hand, standard dialogs and interaction mechanisms can be used, like in familiar graphical user interfaces (GUI). Still, on the other hand, it is also possible to offer direct manipulation of VR content.

1.1. Natural User Interfaces

Natural user interfaces are interfaces which enable users to interact with the system the same way they do in a physical world [15]. Van Dam [15] states that "the ideal user interface would let us perform our tasks without being aware of the interface as the intermediary" (p. 50). Reality-based interfaces (RBI) are interfaces, which try to mimic the real world and allow direct manipulation. Jacob [16] states that "Direct Manipulation moved interfaces closer to real-world interaction by allowing users to directly manipulate objects rather than instructing the computer to do so by typing commands." Therefore, RBI has four different themes: naïve physics, body awareness and skills, environment awareness and skills, and social awareness and skills [16].

- Naïve physics (NP) describes simple physical behavior like gravity, friction, speed, and adhesion, as well as scaling.
- Body awareness and skills (BAS) addresses the point that users have a body independent of the environment.

- Environment awareness and skills (EAS) describe the fact that humans have a physical presence in a 3D environment, which includes objects and landscapes. They increase the spacial orientation.
- Social awareness and skills (SAS) describes that humans are aware of the presence of other humans that can interact with other people socially. These interactions include verbal and non-verbal communication.

Since not every user interface can address each theme most realistically, designers should "give up reality only explicitly and only in return for other desired qualities" [16]. Bowman [17] describe approaches in which "the user's hand becomes the input device—it is not necessary to hold another device in the hand." They propose a system, which enables users to navigate through a menu to define actions to manipulate the virtual environment. Furthermore, they introduce a use case that demonstrates how to use a virtual keyboard for text input, as well as a two-handed virtual-navigation technique.

Human system interaction in Virtual Reality enables users to use a remarkable tool, the human hand. "It allows us to manipulate physical objects quickly and precisely" [18] (p. 252). The parameter of quick interaction influences the usability dimension of efficiency [19]. Therefore, it is one focused factor in the interaction development of VR user interfaces. Another influencing parameter for efficient dialogues is learnability [20]."The interactive system should allow the user to perform the task with minimal learning" [20] (p. 13). Natural user interfaces, therefore, have the advantage that, if they are used to operate software dialogs, they enable users to learn the application with no effort.

In the use case of an assembly task, workers use their hands and fingers to grab and manipulate parts to assemble a product. Transferring these actions into VR would enable a more intuitive use of DHMs. Therefore, we propose a system with an interaction method, which eliminates the mouse and keyboard. The system allows users to rely on virtual representations of their hands in VR to instruct the DHM IMMA. This, in turn, eliminates a human system interaction to define an assembly sequence with the DHM and replaces mouse and keyboard. Instead, it tracks the hands, behavior, and interactions of users in VR. The tracking of human behavior and movements in the context of assembly is a well investigated topic. Bleser et al. [21], e.g., postulate an on-body sensor network to capture user and workspace characterization of industrial workflows.

1.2. Problem Statement

As shown above, it is necessary to use digital human models in virtual product development to assure human factors such as ergonomics. These simulations, however, involve a lot of manual configuration effort using WIMP interfaces. On the other hand, there are reality-based interfaces, which allow a situation, in this case, assembly activities, to be simulated. But the instruction of this DHM is still complicated. Users still have to perform up to 55 actions to instruct IMMA to grab an object with two hands and move it to another place [22].

However, to this day, it is not possible to interact with the virtual scene and use the interactions as input for the manikin. Thus, it is only possible in VR to review the model and the simulation results. The authors thereby transfer the idea of user and environment tracking into the virtual action tracking system (VATS) and create a natural user interface, which allows direct manipulation. Our approach minimizes the necessary menu items, which means that users do not have to navigate through menus in the ergonomic simulation software. This enables them to manipulate the virtual environment directly and thus to define the input for the DHM. We want to use the approach of RBI in our system to provide an effective way to work with DHMs and substitute mouse and keyboard with an interactive VR system. Our goal is to develop a user-friendly application to instruct a digital human model that significantly improves the control of the digital systems' functionality through intuitive human interaction features. We formulated the following hypotheses:

Hypothesis 1. *The time to define an assembly sequence with the VATS is significantly faster than with the existing WIMP Interface.*

Hypothesis 2. *The time to learn to define an assembly sequence with the VATS is significantly faster than with the existing WIMP Interface.*

Hypothesis 3. *The subjective workload with the VATS is less than with the WIMP interface.*

Another exploratory question is to what extent users make mistakes when using the two different applications. An error, using the VATS, is when a user selects a wrong grip type. In the WIMP interface, however, the question is whether the user can complete the complex handle definition at all. Comparable error analyses are not possible, because on the one hand, VATS is always terminated automatically, and on the other hand, the instructions for the WIMP interface specify the grip type. Nevertheless, the two types of errors are illustrated descriptively.

2. Materials and Methods

In this section, we introduce the VATS (virtual action tracking system) and describe the conducted user study.

2.1. Virtual Action Tracking System (VATS)

The basic idea of the VAT system is that all actions performed in VR, which are related to a person or object, are automatically recorded. This recording includes: the position a user is standing in, the line of sight, the interactions (grips) the user is performing, the information about the object the user is interacting with, and the changes in the scene based on the interactions the user performs. The software transfers these action parameters to the DHM. They thus serve as the input for the configuration of the assembly simulation. The DHM now mirrors the interactions of the user with the virtual environment. We have developed four application parts to implement this approach: a CAD-data interface, a user-system interaction module, and a grip recognition system, as well as a data interface to the DHM IMMA (Figure 1).
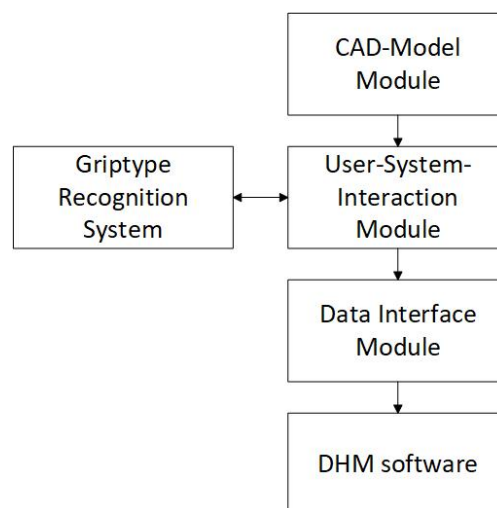

Figure 1. Virtual action tracking system architecture.

2.2. Technical System

The VATS is implemented in Unity and uses the head-mounted display HTC Vive (Figure 2) for immersive visualization and a Leap Motion [23] to track hands and fingers. The Leap Motion is an optical tracking device, which enables users to bring their real hands into the Virtual Environment (VE) and thus use the hands to interact with 3D models. Unity, as an authoring tool, provides the interfaces to the HTC Vive, as well as the Leap Motion tracking system.

Figure 2. HTC Vive with a Leap Motion [24].

2.2.1. CAD-Data Interface

The CAD-data interface allows the import of CAD-data in the data exchange format JT [25]. This enables users to change the environment on which they want to conduct an ergonomic assessment. The virtual product parts and the work station structure is required data input.

2.2.2. User-System-Interaction Module

The module contains the functions that enable users to navigate themselves in the VE as well as select and manipulate 3D models. By walking around, users can navigate freely through the VE. The Leap Motion sensor transfers the real hand into the VE. Figure 3 shows that the Leap Motion sensor is an optical tracking sensor based on infrared light (IR). It contains LED lights and a stereoscopic camera, that captures the hand and finger movements. It has a 120° field of view on the long side and 150° on the short side of the sensor (Figure 3). It is capable of tracking hands with a maximum distance of 80 cm, but the higher the distance, the less accurate the tracking gets [26]. Users do not have to look at their hands due to the vast field of view. Still, their hands need to be in the tracking volume of the leap motion.
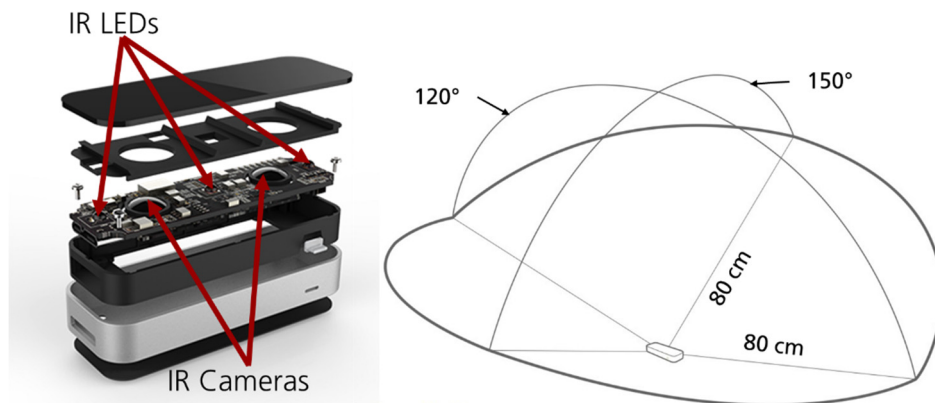

Figure 3. Hardware (left) and the field of view (right) of the Leap Motion Sensor (pictures modified from [27]).

We implemented a simple virtual hand [18], which means that the hand position, in reality, is the same as in VR. According to RBIs, we address EAS, since users can pick up, place, alter, and arrange objects by grasping them with the virtual representation of their hands. In NP, we address the topics: persistence of objects, velocity, and relative scale in the VE. Through the virtual representation of the users' hands, we achieve body awareness in the VE. A higher NP would be achievable by using haptic data gloves. However, as our approach addresses a use case in which users quickly want to use the system, we decided not to use wearable devices.

The system continually measures the distances between the fingers and palm to the surface of the virtual objects (Figure 4).
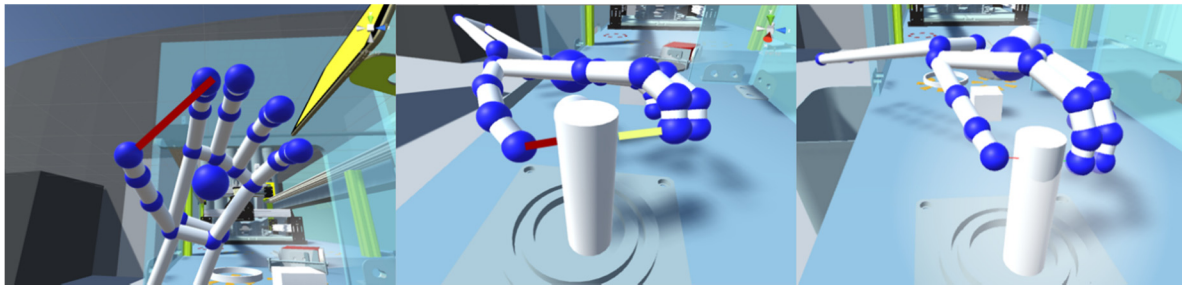
Figure 4. Visualized distance measurements between the hand and the virtual object.

We implemented a second-hand grasping metaphor. There is one leading hand and one following hand. As soon as the leading hand or one finger gets in contact with a virtual object, this part of the hand stays on the surface of the object. If users close their hands even more, a second hand is displayed. This hand represents the position of the real hand. Additionally, the user gets a visual feedback on how far away their fingers are from the surface.

This feedback is provided if the distance between the finger and the surface object is more than 1.5 cm (yellow) or less than 1.5 cm (green) [28]. The virtual representation of the fingers changes color, depending on their distance to the object. After a grip is detected, the users get a user interface with three different grip types. These grip types are the most likely ones of the grip type recognition system (Section 2.2.3). The users select a grip type by moving their hand to the left, to choose the first grip type, to the front to select the second grip type, and to the right to choose the third grip type. After the grasping phase, the object is attached to the hand, and users can move the object freely in the VE. To release the object, the users have to open their hands.

2.2.3. Grip Type Recognition System

IPS IMMA has a preconfigured database with nine different grip types: tip pinch, chuck grip, lateral pinch, pistol grip, prismatic four-finger pinch, spherical grip, cylindrical power grip, diagonal power grip and parallel extension (Figure 5). Engineers use the grip types in IPS IMMA and attach a selected grip from the database to the object which is used for the ergonomic assessment. To achieve the same results with the WIMP interface and the VATS, the VATS needs to detect the grip types of the database. If the system were not able to distinguish between the grip types and use the finger angles to define how the DHM grabs the object, it would be too complicated, almost impossible, for engineers to reproduce the same simulation results. Engineers would then have to define each finger angle independently. This makes it necessary to detect the nine different grip types from IPS IMMA.
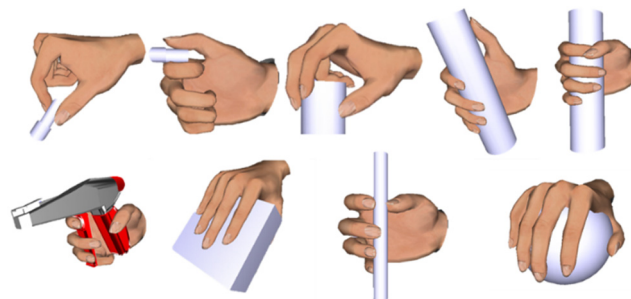

Figure 5. The implemented grip types of the grip type recognition system.

The backend of the grip type recognition system is SciKit Learn [29]. It provides several machine-learning algorithms and can easily be integrated into other software. As we already know the categories (grip types), we implemented a supervised classification algorithm. To identify the appropriate algorithm, we recorded between 120 and 400 datasets for each grip type.

Two expert users used the leap motion to grab the objects with the nine different grip types. Figure 6 presents the objects. One dataset contains the following values:

- There were 14 joint angles: angles of one hand: two for the thumb, three for index, middle, ring, and pinky;
- Four finger distances: distances between thumb to index finger; index to the middle finger, middle finger to ring finger and ring finger to pinky finger;
- Four thumb distances between the thumb and the index, middle, ring and pinky finger;
- Two values for the orientation of the hand;
- Six binary values: information if a finger or the palm is in contact with the surface.
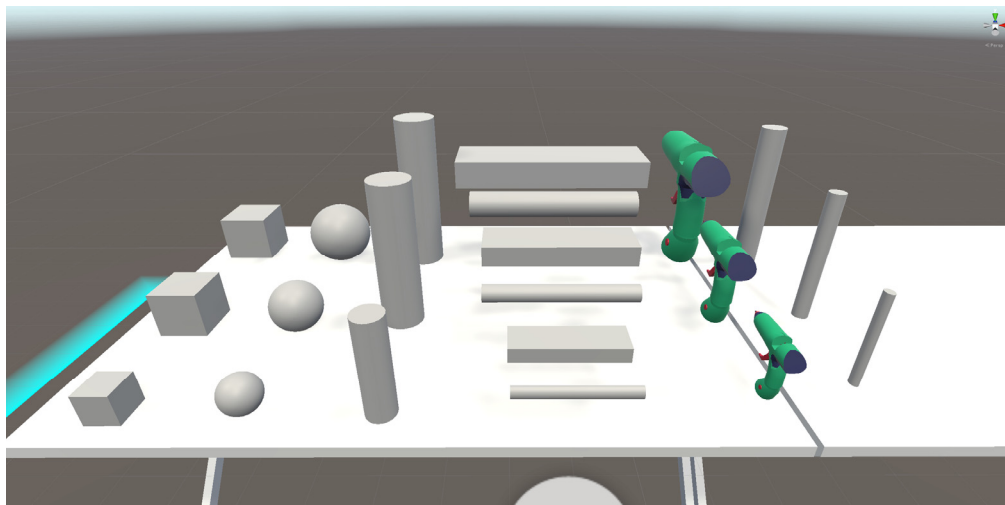

Figure 6. Training scene with the objects for the machine learning algorithms. Three dice, thee balls, three horizontal cylindric objects, three vertical cylindric objects, three cuboids, and three drilling machines, as well as three smaller horizontal cylindric objects.

We split the data into 70% training data and 30% testing data and compared the correct prediction between the random forest classifier (0.98), nearest neighbors' classifier (0.93), as well as the support vector classification (0.96). These results show that the random forest classifier performs best for this application. The grip type recognition system analyzes the features and sends back the highest-ranked grip types. Users intend to define a specific grip type. If the system recognizes the wrong grip type, the manipulation in VR still works. Still, to be able to conduct the ergonomic simulation, the user then has to perform the task again with the correct grip type. The recognized grip type is used as the input for the DHM IMMA for ergonomic simulations.

2.2.4. DHM Data Interface

The DHM data interface is the interface between the VE and the DHM software IPS IMMA 3.4.3. Thus, the following information is exported from the VE: object: name, start, and endpoint of the manipulation and the path; grip: grip type, grip point including position and orientation, percentage of how closed the hand is. This information, in combination with the interaction sequence, is sufficient to instruct the DHM IMMA.

2.3. User Study

2.3.1. Participants

A sample of N = 29 persons participated in the study; 21 already had experience with VR. Eight participants furthermore had experience with a Leap Motion and five with Digital Human Models

(Siemens Jack and Dassault HUMAN). The average age was 28.8 years. Eighteen were students, nine were researchers at the Fraunhofer IPK, one informatics, and one IT-specialist.

2.3.2. Measures

In this study, we captured objective, as well as subjective, criteria. To measure the subjective workload of defining an assembly sequence for a digital human model with the WIMP and the VATS, we used the NASA-TLX questionnaire [30]. The questionnaire addresses the following dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. With this questionnaire, we want to investigate the subjective workload users have when using the different user interfaces to define the assembly sequence. Furthermore, to investigate the learnability, we use the corresponding part of the standardized questionnaire for dialogue design [31,32]. The questionnaire is based on the ISONORM 9241/110 and has seven dimensions. The dimension we focused on was "suitability for learning." This dimension contains five items: the time required to conduct the task, encouragement users have to test it, and the memorization of details, memorability, and learning without help. The average of these is the value for suitability for learning. The scale is between — until +++. Additionally, we created a questionnaire with ten questions. The aim was to get feedback on which system (VR vs. WIMP) the participants liked more, and if they would use the VATS and why. Two questions also focused on technical issues, like if the participants can imagine using gloves instead of a Leap Motion and using an HMD at their daily workplace. The scale ranged from 1 (do not agree at all) to 6 (fully agree).

As objective data, we also measured two different values. The task completion time (TCT) and the amount of errors users did during the task.

2.3.3. Study Environment

The hardware setup for the study consisted of a Dell Alienware 15 R3 Laptop (Intel Core I7 6820HQ, 16 GB of RAM, Nvidia Geforce 1070 with 8GB GDDR5 RAM). We used an HTC Vive Business Edition, including two SteamVR base stations for room-scale tracking. The software was always running with 90fps and was implemented in Unity 3D 5.6. The use case was a pick and place task. This task is a typical task that exists in a lot of factories. A worker must pick a specific part from a box or shelf and put it onto a factory line. It is straightforward to conduct this task in reality, but analyzing this task (grab, move, release) with a digital human model is already challenging enough to suit as a task for this study. The scene displayed a 3D model in a box that had to be moved to a particular position on an assembly line. The assembly line was in a static place. The interactions users had to perform with the VATS are descibed in Figure 7. In Figure 8 we describe the actions a user performs to define the assembly sequence with the WIMP interface of IPS IMMA.
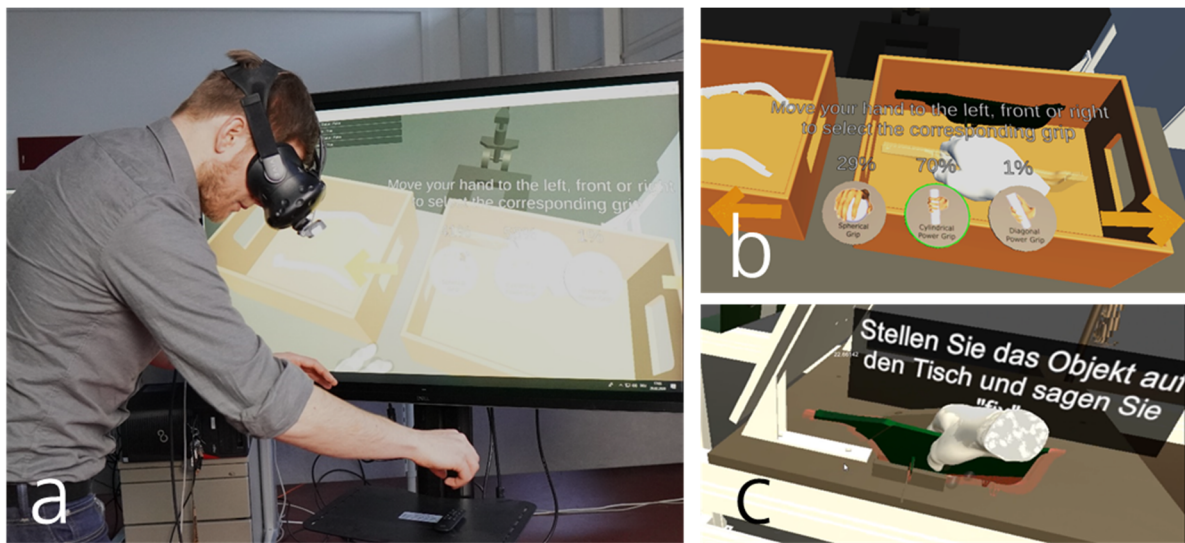
Figure 7. User interaction with the VATS (**a**) user with the head-mounted display and leap motion. He is grabbing the virtual object. (**b**) Grabbing an object and visualization of the three most likely grip types. By moving the hand to the left, front, and right it is possible to select the left, the middle, or the right grip type. (**c**) The user places the object at the final position on the assembly line.
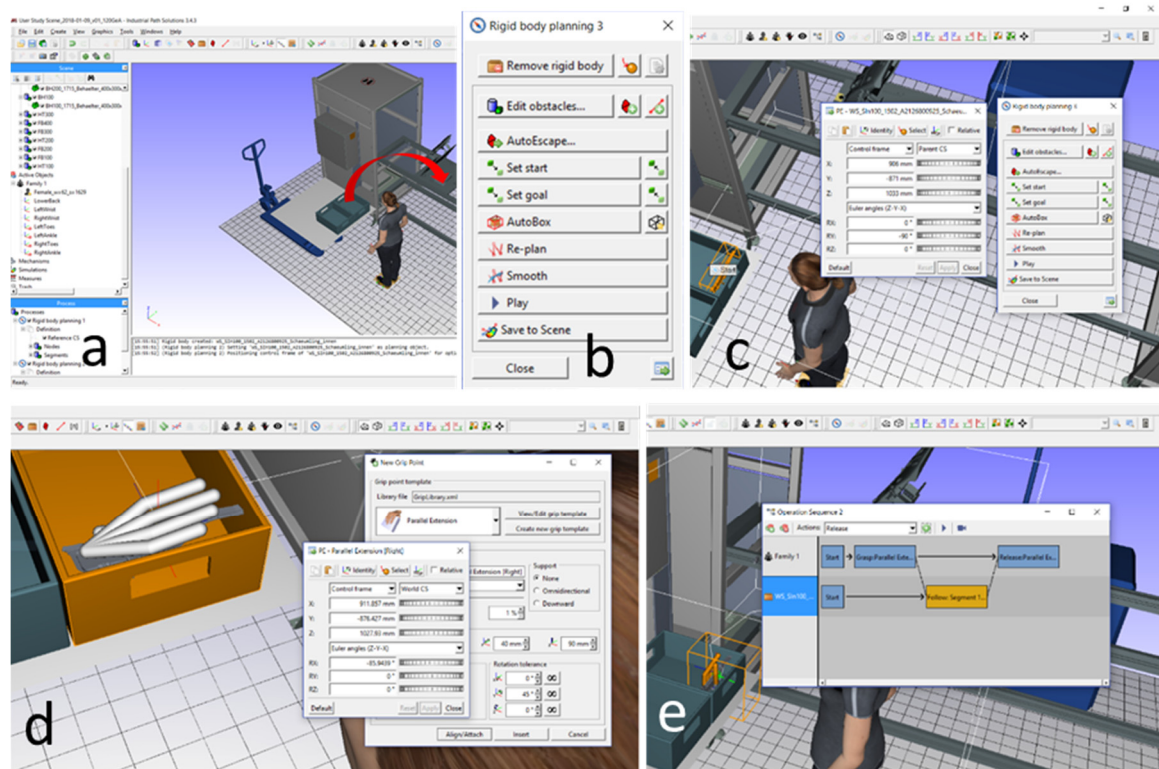

Figure 8. (**a**) Overall scene of the task. (**b**) The user menu to define the start-point, end-point, as well as the collision-free path. (**c**) Example of the user interface to define a specific points, with sliders for each axis and orientation. (**d**) Example of the interface to define the grip. (**e**) Operation sequence editor to define the sequence of the task—loading the object, which is moved in the task—loading the manikin family—by drag and drop to define the sequence (grab the object, move it along the path, release the object).

2.3.4. Procedure

At the beginning of the study, the participants had to fill out a demographic questionnaire. Fifteen started with the VATS, and 14 with the IPS IMMA WIMP interface. In VR, they conducted the task 9 times. After that, they had to do the pick and place task with the WIMP interface. Before the start of the study, they got an accurate description of the task they had to perform. In VR, the users had to look at a green cross for three seconds to start the task, and the time measurement started as well. Then, they had to grab the object, select one of the three presented grip types, and put the gabbed object onto the assembly line, and release it. The time measurement stopped when the participants released the object. From the available nine grip types, the users had the task to either use the spherical or parallel grip. As we want to evaluate the workload users have using the different user interfaces (VATS vs. WIMP), they had to fill out the NASA-TLX. After finishing all trials, the participants had to fill out the learnability questionnaire. Then, they switched to the IPS IMMA WIMP interface. The process was the same as in VR. The participants got an accurate description of the task. Then, they had to perform the task eight times and had to fill out the NASA-TLX after each trial. After that, they had to fill out the learnability questionnaire. At the end of the study, the participants furthermore filled out the final questionnaire.

3. Results

In this section, we present the results of the user study. We divided the results into objective and subjective measures.

3.1. Objective Measures

Table 1 shows the results of the task vompletion time (TCT) for the WIMP and VATS. The WIMP interface has the highest value of 435 s in the first trial and the lowest in the eighth trial (128 s). Individual *t*-tests were calculated between each trial. No ANOVA was calculated for all data, since only the main effects, i.e., general differences between VATS and WIMP and the eight trials, can be determined with the ANOVA. For us, the differences between the trials within one system (WIMP or VATS) are relevant. Since each data set was used twice, the *p*-value was corrected to 0.025. With the WIMP interface, there are significant differences between the trials one until four and significant differences between the fifth until the seventh trial. The VATS interface has a TCT of 59 s in the first and 23 s in the eighth trial. There only exist significant results between the first until the third trial. After that, no significant differences occur (Table 1). Besides, Figure 9 shows the errors for the WIMP interface and the VATS. An error in the VATS was counted when a user did not select the right grip type to perform the task. An error with the WIMP interface was counted when users were not able to fulfill the task.

Table 1. Test results for eight trials for the VATS as well as the WIMP. In the VATS, nine trials were conducted. Thus, we also tested the trial eight vs. trial nine. The *p* values with a grey background indicate the significant results of the *t*-test between trial R(X) and trial R(X + 1).

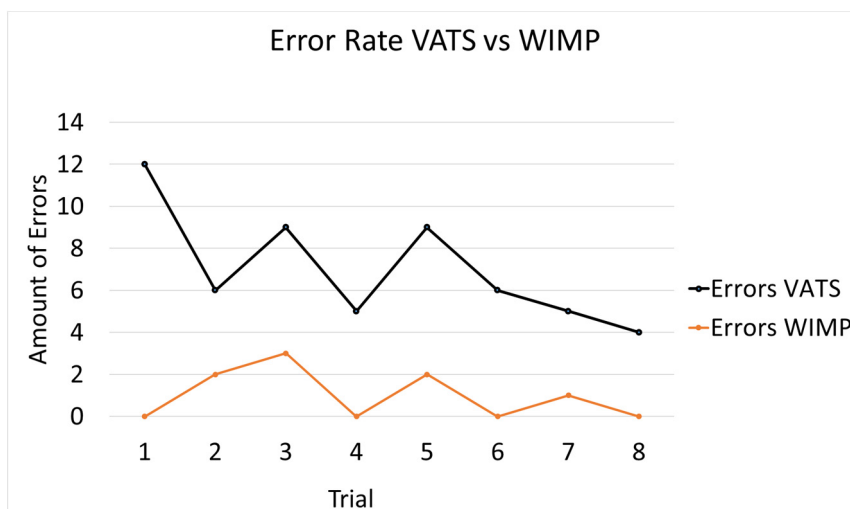|	R1	R2	R3	R4	R5	R6	R7	R8
mean TCT VATS	59 s	35 s	25 s	27 s	24 s	25 s	24 s	23 s
p (*t*-test R/R + 1) VATS	0.001	0.018	0.601	0.582	0.769	0.867	0.735	0.186
mean TCT WIMP	435 s	281 s	210 s	179 s	166 s	142 s	132 s	128 s
p (*t*-test R/R + 1) WIMP	<0.01	<0.01	<0.01	0.193	<0.01	0,02	0.314	

Figure 9. Amount of errors for VATS and WIMP. VATS error = wrong selected grip type. Error WIMP = not completed task.

Additionally, Figure 9 shows the errors for the WIMP interface and the VATS. On average, seven out of 29 users made errors with the VATS in one trial. However, the errors decrease throughout use. This means that even during repeated use, users selected incorrect grip types. With the WIMP application, an error was counted when the task was not completed correctly. The average amount of errors was one.

3.2. Subjective Measures

To analyze the learnability, we used the previously described learnability questionnaire. The results show that the VATS has a significantly higher learnability than the WIMP interface. Whereas the VATS has positive values between 1.0 (functions) and 2.0 (learning), the WIMP has values from −0.5 (learning) until −1.8 (details), (Figure 10).
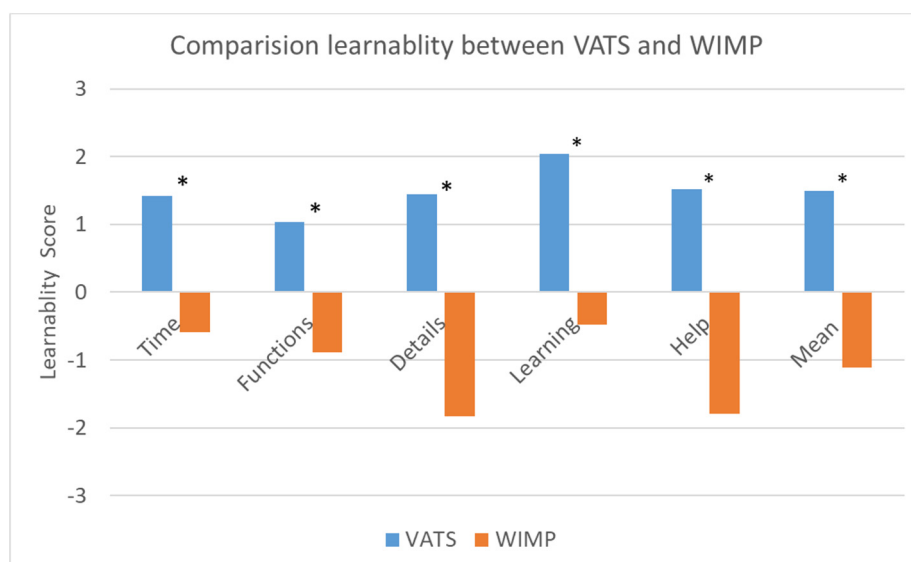

Figure 10. Results of the learnability questionnaire. The values are (first value VATS, second WIMP): Time: 1.4/−0.6; Functions: 1.0/−0.9; Details: 1.4/−1.8; Learning: 2.0/−0.5; Help: 1.5/−1.8; Mean: 1.5/−1.1

Figure 11 shows the results for each dimension of the NASA-TLX. We present the mean values of all participants for the trials, one until eight. The results show a significant higher temporal demand

between the VATS (17.5) and WIMP (32.2) interface. The other five, as well as the mean overall the dimensions, do not show any significant differences.
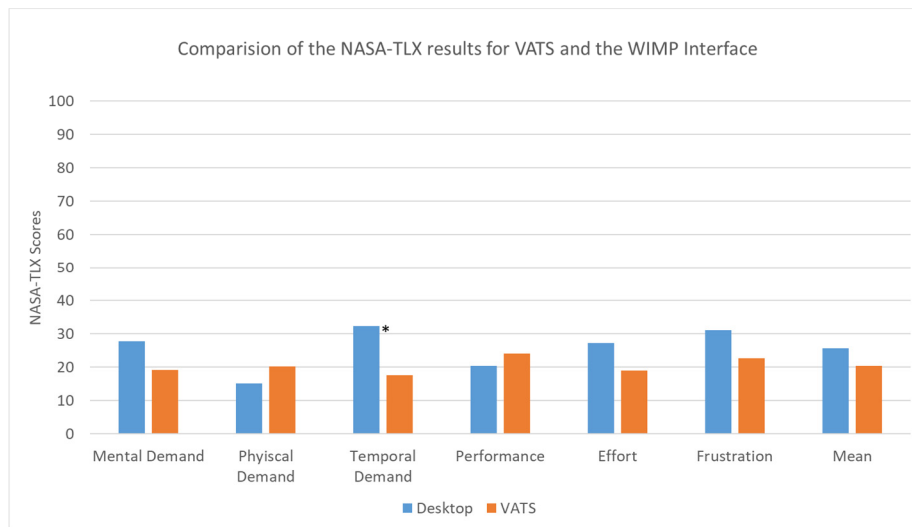

Figure 11. Results of the NASA-TLX for each dimension.

In Table 2, we present the results of the final questionnaire. We conducted a *t*-test against the value of 3.5, as 3.5 is the mean of the scale between 1 and 7. This value is chosen as a test value, because it corresponds to the neutral statement (neither positive nor negative) in the verbal anchoring.

Table 2. Results of the final questionnaire about the VATS and the results of the *t*-test between the mean value of the results and the mean value of the scale. The *p* values with a grey background indicate the significant results of the *t*-test.

	Mean Values	*p* (*T*-Test Mean Values/Mean of Questionnaire Scale)
Improves Interface	4.34	<0.01
Higher Efficiency	4.69	<0.01
Less time consuming	5.41	<0.01
Higher intuitiveness	5.17	<0.01
Increases the use	4.66	<0.01
Sequence definition is easier	5.1	<0.01
System accuracy	3.41	0.77
System supports user	5.17	<0.01
Tracking is precise enough	3	0.06

4. Discussion

The TCT results show that using the VATS, with its direct manipulation in a virtual environment, is significantly faster than using the existing IPS IMMA WIMP interface. This coincides with the work of [33], which states that when you implement a direct manipulation interface, users can transfer their already learnt skills quickly to the new interface. Thus, a well-implemented direct manipulation interface achieves, in this use case, better results. Additionally, the VATS provides only the information which is necessary to conduct the specific task. The user interface of IPS IMMA, as a user interface for experts, offers additional features, that are not necessary, to do the examined use case. In the VATS, on the interaction side, we only implemented grabbing, moving, and releasing objects. The

WIMP interface of IPS IMMA, e.g., allows for unconstrained certain degrees of freedom when grasping an object or using one hand as a support hand. Furthermore, the placement of grip positions on a 2D desktop is a challenge for every user, including experts, as they must switch between different perspectives to place the grip type on the right position and also adjust the orientation. However, by being able to grasp the object with the own hand in an immersive virtual environment, this problem is eliminated. Thus, our hypothesis that users can conduct the task faster with the VATS can be accepted.

On the other hand, the error rate with the VATS is higher than with the WIMP interface. We expect this for two reasons. For the WIMP interface, the participants had a precise manual of which actions they had to perform to conduct the task. They also had the opportunity to check this manual while doing the task. On the other hand, we presented the users a precise description of how to use the VATS before they put on the head-mounted display. Thus, they could not look up the instructions again while using the VATS. Additionally, none of the participants used the VATS before and knew how the interaction works, while users are highly familiar with WIMP interfaces and using mouse and keyboard in general. But, the amount of errors declined from 12 in the first round until four in the eighth round. As we create the same information with the VATS as with the WIMP, it is still possible to make smaller adjustments, like changing the grip type with the WIMP interface. This hybrid approach enables users to eliminate the errors.

The VATS also enables users to learn to define an assembly sequence faster. The results show that the participants only needed three trials to learn to define the assembly sequence with the VATS. There were no significant improvements in time after the third trial. Using the WIMP interface, the participants had significant improvements in the TCT until the seventh trial. The questionnaire for the suitability of learning also shows these results. In each dimension, the VATS has a significantly better score. These results are also supported by the work of [15], as a user interface with direct manipulation metaphors provides interaction techniques that users know from the real world and can thus need lees trials to learn the interaction process. To sum up, the VATS is easier to learn, and we can accept hypothesis two as well.

The subjective workload was measured with the NASA-TLX questionnaire. The results show that the workload has no significant differences between the items, except in the dimension time. The participants experienced a higher pressure of time using the WIMP interface. This, in our opinion, comes from the number of different tasks the users have to perform to define the assembly sequence in the IPS IMMA (WIMP) interface. Thus, the users needed significantly more time to conduct the task of defining the assembly sequence with the WIMP interface than with the VATS. This leads to the point that the VATS creates less temporal workload, and we have to decline hypothesis three, as we expected no difference between the VATS and the IPS IMMA interface.

The self-developed questionnaire shows significant positive results, except in the item "system accuracy" and that the "tracking system is accurate enough." In our system, we use the leap motion as the tracking device. Whereas a previous study shows [34] that the system accuracy was significantly negative, we implemented several features to improve the accuracy: we rotated the leap motion at an angle of 45° so that the optimal tracking volume is in front of the users' torso. Furthermore, we implemented a warning for the users, which tells them if the hand has a higher distance than 45 cm between the leap motion and the wrist. Thus, the users always had their hands in the optimal tracking volume.

To sum up, the VATS is a system that enables engineers to define an assembly sequence. The time to conduct the definition of the assembly sequence is lower than with the existing WIMP interface. Still, it is not as precise as the WIMP interface, but users can correct the errors from the VATS in the WIMP interface.

In general, we expect, after reducing the error rate of the VATS, that more companies and engineers are going to use digital human models, as the engineers won't need the same amount of training, due to the more natural way to define an assembly sequence. This can improve the ergonomics of many assembly stations. In the end, this can have a positive influence on the workers who have to

conduct the assembly tasks in the real factory lines daily. We don't see any impediments of using a head-mounted display, as our participants mentioned they would use head-mounted displays daily at their workplace.

Future work will focus on improving the interaction method. We are going to work on using haptic gloves to interact in the virtual environment and use the reaction forces of each finger, as well as the interactions as simulation inputs, for the assembly sequence. In addition to this, we will extend the system to be able to interact with flexible parts like cables and flat flexible parts (e.g., car door panels). Furthermore, we are going to investigate different hand and finger tracking devices to improve hand tracking accuracy. We expect to be able to use the VATS also for other DHMs like JACK, Human Builder, Ramsis, and EMA, and will investigate the possible necessary adjustments to the VATS.

Furthermore, we see several different use cases for the system:

- Creation and validation of training for assembly and MRO (maintenance, repair, and overhaul);
- Creation and validation of assembly and MRO assistance systems, e.g., Augmented Reality, digital worktables, adaptive instructions;
- Creation and validation of fitting and assembly processes itself, e.g., collision-free assembly planning.

These use cases will be addressed in further research projects.

Author Contributions: Conceptualization, A.G.; Methodology, A.G. and E.B.; Software, A.G.; Validation, A.G., E.B.; Resources, R.S.; Writing—original draft preparation, A.G. and E.B.; writing—review and editing, A.G., E.B., and R.S.; visualization, A.G.; supervision, R.S.; project administration, R.S.; funding acquisition, R.S. All authors have read and agreed to the published version of the manuscript.

Ethics Statement: In our institute, it is not mandatory to prove an ethic statement. Nevertheless, we have oriented ourselves to the guidelines of the German Society of Psychology (DGP). Upon starting the study, the participants were briefed orally about the study so that they could understand what we want to do. Additionally, all participants signed a consent form that included all the following points: the participants were informed that data collected is handled confidential and processed anonymously. They signed that they were informed about possible occurring simulator-sickness, and if it occurs, they should pause or cancel the experiment without adverse effects. This consent, obtained from all participants, was both written and informed.

References

1. Anderl, R.; Eigner, M.; Sendler, U.; Stark, R. *Smart Engineering. Interdisziplinäre Produktentstehung*; Springer: Berlin, Germany, 2012; ISBN 978-3-642-29371-9.
2. Bubb, H.; Bengler, K.; Grünen, R.E.; Vollrath, M. *Automobilergonomie*; Springer Fachmedien: Wiesbaden, Germany, 2015; ISBN 9783834822970.
3. Hs Group. RAMSIS Automotive—Effiziente Innenraumentwicklung—Ohne Nachbessern. Available online: https://www.human-solutions.com/mobility/front_content.php?idcat=252&lang=3 (accessed on 12 March 2020).
4. Dassault Systèmes®. 3D Design & Engineering Software. Available online: https://www.3ds.com/ (accessed on 12 March 2020).
5. Spanner-Ulmer, B.; Mühlstedt, J. Digitale Menschmodelle als Werkzeuge virtueller Ergonomie. *Ind. Manag.* **2010**, 69–72. (In Deutsch)
6. Wischniewski, S. *Digitale Ergonomie 2025. Trends und Strategien zur Gestaltung Gebrauchstauglicher Produkte und Sicherer, Gesunder und Wettbewerbsfähiger Sozio-Technischer Arbeitssysteme Forschung Projekt F 2313; [Abschlussbericht]*; Bundesanstalt für Arbeitsschutz und Arbeitsmedizin: Berlin, Germany, 2013; ISBN 3882610042.
7. Hanson, L.; Högberg, L.; Bohlin, R.; Carlson, J.S. IMMA—Intelligently Moving Manikins—Project Status 2011. In Proceedings of the First International Symposium on Digital Human Modeling, Lyon, France, 14–16 June 2011; ISBN 2953951504.

8. Bertilsson, E.; Hanson, L.; Högberg, D. Creation of the IMMA manikin with consideration of anthropometric diversity. In Proceedings of the 21st International Conference on Production Research, Stuttgart, Germany, 31 July–4 August 2011; pp. 416–420, ISBN 978-1-63439-826-8.

9. Arnaldi, B.; Guitton, P.; Moreau, G. *Virtual Reality and Augmented Reality: Myths and Realities*; Wiley: Hoboken, NJ, USA, 2018; ISBN 9781119341048.

10. Cruz-Neira, C.; Sandin, D.J.; DeFanti, T.A. Surround-screen projection-based virtual reality. In Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, Anaheim, CA, USA, 2–6 August 1993; pp. 135–142.

11. Brandenburg, E. VR-Twin: Monteurgerechte Anlagenumplanung. In *FUTUR: Vision Innovation Realisierung: Vernetzung in der Produktion*; Uhlmann, E., Ed.; Fraunhofer IPK: Berlin, Germany; p. 12.

12. Dittrich, E.; Israel, J.H. Arrangement of Product Data in CAVE Systems. Available online: http://web.cse. ohio-state.edu/~{}raghu/teaching/CSE5544/Visweek2012/scivis/posters/ (accessed on 27 February 2020).

13. middleVR. VR Teleconferencing Software For CAD Teams–Improov, Collaboration in VR. Available online: https://www.improovr.com/home-v2/ (accessed on 27 February 2020).

14. ESI Group. IC.IDO|ESI Group. Available online: https://www.esi-group.com/de/software-loesungen/virtual-reality/icido (accessed on 27 February 2020).

15. van Dam, A. Beyond WIMP. *IEEE Comput. Grap. Appl.* **2000**, *20*, 50–51. [CrossRef]

16. Jacob, R.J.K.; Girouard, A.; Hirshfield, L.M.; Horn, M.S.; Shaer, O.; Solovey, E.T.; Zigelbaum, J. Reality-based interaction: A Framework for Post-WIMP Interfaces. Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems—CHI '08, Florence, Italy, 5–10 April 2008; ISBN 9781605580111.

17. Using Pinch Gloves for both Natural and Abstract Interaction Techniques in Virtual Environments. Available online: http://eprints.cs.vt.edu/archive/00000547/ (accessed on 12 March 2020).

18. LaViola, J.J.; Kruijff, E.; McMahan, R.P.; Bowman, D.A.; Poupyrev, I. *3D User Interfaces. Theory and Practice*; Addison-Wesley: London, UK, 2017; ISBN 9780134034324.

19. *Ergonomie der Mensch-System-Interaktion Teil 11: Anforderungen an die Gebrauchstauglichkeit—Leitsätze*; Deutsches Institut für Normung e.V.: Berlin, Germany, 1999.

20. DIN. *Ergonomie der Mensch-System-Interaktion Teil 110: Grundsätze der Dialoggestaltung*; DIN: Berlin, Germany, 2008.

21. Bleser, G.; Damen, D.; Behera, A.; Hendeby, G.; Mura, K.; Miezal, M.; Gee, A.; Petersen, N.; Maçães, G.; Domingues, H.; et al. Cognitive Learning, Monitoring and Assistance of Industrial Workflows Using Egocentric Sensor Networks. *PLoS ONE* **2015**, *10*, e0127769. [CrossRef] [PubMed]

22. Brandenburg, E.; Geiger, A.; Rothenburg, U.; Stark, R. Anforderungsanalyse einer Montagesimulationssoftware zwecks Aufwandsminimierung manueller Konfigurationsaufgaben. In *Useware 2016: Mensch-Technik-Interaktion im Industrie 4.0 Zeitalter*; VDI-Verlag: Düsseldorf, Germany, 2016; pp. 15–23. ISBN 978-3-18-092271-3.

23. Leap Motion. Available online: https://www.leapmotion.com/ (accessed on 21 May 2018).

24. Colgan, A. Leap Motion and HTC Vive FAQ: What You Need to Know. Available online: http://blog. leapmotion.com/leap-motion-htc-vive-faq/ (accessed on 27 February 2020).

25. International Organization for Standardization. *Industrial Automation Systems and Integration—JT File Format Specification for 3D Visualization*, 2nd ed.; ISO: Geneva, Switzerland, 2017.

26. Weichert, F.; Bachmann, D.; Rudak, B.; Fisseler, D. Analysis of the accuracy and robustness of the leap motion controller. *Sensors* **2013**, *13*, 6380–6393. [CrossRef] [PubMed]

27. Colgan, A. How Does the Leap Motion Controller Work? Available online: http://blog.leapmotion.com/ hardware-to-software-how-does-the-leap-motion-controller-work/ (accessed on 11 March 2020).

28. Geiger, A.; Bewersdorf, I.; Brandenburg, E.; Stark, R. Visual feedback for grasping in virtual reality environments for an interface to instruct digital human models. In Proceedings of the AHFE 2017 International Conference on Usability and User Experience, Los Angeles, CA, USA, 17–21 July 2017; ISBN 9783319604916.

29. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

30. Hart, S.G.; Staveland, L.E. *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*; Elsevier: Amsterdam, The Netherlands, 1988; pp. 139–183. ISBN 9780444703880.

31. Prümper, J.; Anft, M. Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung—Ein Fallbeispiel. In *Software-Ergonomie '93*; Rödiger, K.-H., Ed.; Vieweg + Teubner Verlag: Wiesbaden, Germany, 1993; pp. 145–156. ISBN 978-3-519-02680-8.

32. Prümper, J. Test IT: ISONORM 9241/10. In Proceedings of the HCI International '99 (the 8th International Conference on Human-Computer Interaction), Munich, Germany, 22–26 August 1999; pp. 1028–1032, ISBN 0805833919.

33. Wigdor, D.; Wixon, D. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*; Morgan Kaufmann/Elsevier: Amsterdam, The Netherland, 2011; ISBN 978-0-12-382231-4.

34. Geiger, A.; Brandenburg, E.; Stark, R. Virtual User Interaction to Instruct Digital Human Models for Effective Model Based Engineering—An Expert Review. In *Virtual Reality and Augmented Reality*; Barbic, J., D'Cruz, M., Latoschik, M.E., Slater, M., Bourdot, P., Eds.; Springer: Cham, Switzerland, 2017; pp. 13–14. ISBN 978-3-319-72322-8.