**Yaiza Vélaz**
Department of Mechanical Engineering,
CEIT and TECNUN,
University of Navarra, Manuel Lardizabal 15,
San Sebastián 20018, Spain
e-mail: yvelaz@gmail.com

**Jorge Rodríguez Arce**
Department of Mechanical Engineering,
CEIT and TECNUN,
University of Navarra,
Manuel Lardizabal 15,
San Sebastián 20018, Spain
e-mail: jorge.arce.mx@gmail.com

**Teresa Gutiérrez**
Industry and Transport Division,
C/Geldo—Parque Tecnologico de Bizkaia,
Edificio 700,
Derio 48160, Spain;
TECNALIA,
San Sebastián 20009, Spain
e-mail: teresa.gutierrez@tecnalia.com

**Alberto Lozano-Rodero**
Department of Mechanical Engineering,
CEIT and TECNUN,
University of Navarra,
Manuel Lardizabal 15,
San Sebastián 20018, Spain
e-mail: alozanorodero@gmail.com

**Angel Suescun**
Department of Mechanical Engineering,
CEIT and TECNUN,
University of Navarra,
Manuel Lardizabal 15,
San Sebastián 20018, Spain
e-mail: asuescun@ceit.es

# The Influence of Interaction Technology on the Learning of Assembly Tasks Using Virtual Reality

*This paper focuses on the use of virtual reality (VR) systems for teaching industrial assembly tasks and studies the influence of the interaction technology on the learning process. The experiment conducted follows a between-subjects design with 60 participants distributed in five groups. Four groups were trained on the target assembly task with a VR system, but each group used a different interaction technology: mouse-based, Phantom Omni® haptic, and two configurations of the Markerless Motion Capture (Mmocap) system (with 2D or 3D tracking of hands). The fifth group was trained with a video tutorial. A post-training test carried out the day after evaluated performance in the real task. The experiment studies the efficiency and effectiveness of each interaction technology for learning the task, taking in consideration both quantitative measures (such as training time, real task performance, evolution from the virtual task to real one), and qualitative data (user feedback from a questionnaire). Results show that there were no significant differences in the final performance among the five groups. However, users trained under mouse and 2D-tracking Mmocap systems took significantly less training time than the rest of the virtual modalities. This brings out two main outcomes: (1) the perception of collisions using haptics does not increase the learning transfer of procedural tasks demanding low motor skills and (2) Mmocap-based interactions can be valid for training this kind of tasks.* [DOI: 10.1115/1.4028588]

*Keywords: virtual reality, haptic devices, motion capture, learning, assembly training, knowledge transfer*

## Introduction

Applications based on VR are aimed at simulating real or imaginary worlds by means of computer technology. VR has been applied to very different contexts, and we find applications related to entertainment, animation, therapeutic purposes, and training purposes, among others. VR allows extra information that is not available in the real world to be displayed, which turns out to be particularly interesting in learning contexts [1]. This paper focuses on the use of VR systems for learning procedural tasks, specifically learning maintenance and assembly tasks. These complex tasks involve the knowledge of specific procedures and techniques for each machine, which are usually taught through video documentation, 2D mechanical drawings, or the explanation from an expert. However, most researchers agree that procedural tasks are learnt gradually as a result of practice through repeated exposures to the particular task. In this sense, VR systems can provide some benefits with respect to traditional systems as they can support a learning-by-doing approach. Thus, VR systems allow trainees to practice the task as many times as they need, while at the same time eliminating various constraints related to training in real environments. There are only a few research papers that have studied the efficiency of virtual assembly training systems, and the knowledge transfer acquired with them. Those results are limited to a few tasks which do not involve tools manipulation. Therefore, more evaluations of the effectiveness of VR systems for training assembly tasks are needed.

When studying the learning acquired through a specific virtual system, it is interesting not only to study the virtual system as a whole but also to study which factors may influence that learning, i.e., user profile, learning strategy used, type of interaction technology, quality of virtual graphics, and visualization. For example, the kind of feedback or aid given as part of the learning strategy influences the learning process, not only through the quality of the aid (it should be easily interpretable) but also by the quantity (the trainees become increasingly dependent on the guidance, which may inhibit their ability to perform the real task when it is no longer available). Similarly, the type of interaction technology also plays an important role in the acquisition of knowledge. An interaction technology that is difficult or nonintuitive could increase the trainees' cognitive load and reduce their concentration on the assembly task, thus affecting the learning procedure.

This paper focuses on analyzing the influence of the VR system interaction technology in the learning of industrial assembly tasks. Specifically, this paper addresses the following research questions:

Contributed by the Computers and Information Division of ASME for publication in the Journal of Computing and Information Science in Engineering. Manuscript received July 25, 2012; final manuscript received September 7, 2014; published online October 7, 2014. Editor: Bahram Ravani.

- How does the interaction technology affect the learning of a procedural assembly task?
- Which interaction technology is more efficient, i.e., less time consuming, for virtual training?
- Does the interaction technology affect the usability of the virtual training system?

To address these questions, this paper presents a study conducted in order to evaluate the knowledge transfer acquired with different interaction technologies. Based on the work of Yuviler-Gavish et al. [2], the cognitive skill is the main skill required for learning procedural assembly tasks. In order to make our results valid, the participants involved in the experiment had to know the necessary motor skills, so that they could focus on the learning procedure. Results are compared with a traditional training method based on a professional video-tutorial that shows how to perform the task.

For this study, four interaction technologies were selected, which hereafter are shown by categories and the particular device in parentheses:

- Standard PC interaction (2D mouse): This is considered the most common input device for a computer after the keyboard. Commercial applications in the area of assembly simulation mostly involve interacting with this device (i.e., solid edge application).
- Physical and novel interaction (LHIFAM haptic device): Most research contributions propose a haptic device for physical interaction with virtual objects. We can find various types of haptic devices: haptic desktop devices (i.e., PHANToM®), floor-grounded haptics (i.e., LHIFAM), and wiregloves (i.e., CyberGlove® or CyberForce®). In this study, we wanted to use a haptic device that provided high-resolution force feedback and a large workspace in order to work in conditions that are more similar to the real environment. For these reasons we selected the LHIFAM haptic device.
- Natural and novel interaction (3D Markerless Motion capturing [MMocap3D] system and 2D Markerless Motion capturing [Mmocap2D] system): Since commercial and low cost natural interactions entered the entertainment area, this interaction became extremely popular in a very short time. Its robustness has proved to be good enough for a wide range of applications (i.e., therapeutic purposes, military training, or animation). Thus, we implemented and analyzed two systems, MMocap2D and MMocap3D, which differ in the number of cameras used for tracking. The cheapest system, MMocap2D, includes single-camera tracking so we can only track in 2D, which is similar to the mouse condition with the exception that movements are more realistic. EyeToy for PlayStation®2 is an off-the-shelf system similar to ours. The second system, MMocap3D, has a stereo RGB (red, green, and blue) camera from which we obtain 3D hand positions. Analogous low cost commercial systems are Microsoft Kinect™ and Asus Xtion.

The experimental task selected for the study is based on a real assembly task from an industrial company and consists of assembling part of an electronic actuator. It is composed of 23 steps grouped into the following five subtasks: place the two level sensors, the support plate, the electronic board, the actuator cover, and finally the clamp. Following a between-subjects design, 60 participants were randomly assigned to one of the five available training conditions: one traditional training condition and four virtual training conditions (one for each selected interaction technology—Mmocap3D, Mmocap2D, mouse, and haptic). All virtual training conditions used the same VR system; the only difference was in the technology used for the interaction with the virtual objects. One day after training on the task, a post-training test evaluated user performance on the real task.

This paper starts with a review of related work, and then analyzes the advantages and disadvantages of the analyzed interaction technologies. After that we describe the experiment design, results, and conclusions.

## Related Work

We can find contributions to virtual assembly training starting two decades ago. In recent years, one of the most active and promising research topics has been the use of haptic interaction. Pere et al. [3] implemented a computer-based system for virtual assembly training. To interact with the virtual scene, they used the Rutgers Master II exoskeleton haptic device, which tracks hand gestures. Virtual objects could be touched or grasped, and collisions with objects were provided by the haptic device. Brough et al. [4] and Schwartz et al. [5] proposed a virtual environment (VE) for manufacturing tasks called Virtual Training Studio (VTS). The novelty was that they logged users while training in order to give personal guidance if requested. Another example in this area is presented by Bhatti et al. [6]. They proposed a multimodal prototype for virtual assembly training called HIVEx. It was composed of a PHANToM® haptic device, 5DT Data Glove, and stereoscopic displays. Flock of Birds was also used to track the position and orientation of the head mounted display and was synchronized with the virtual view. Abate et al. [7] proposed and implemented an interaction environment in which they simulated maintenance tasks from the aerospace industry. The interaction with the virtual system was accomplished by means of head/hand trackers, CyberGlove, and CyberForce. Seth et al. [8,9] proposed a dual interaction for assembly using PHANToM haptic devices, where users could assemble using both hands simultaneously. They implemented direct object manipulation simulation which allowed the user to grasp and release virtual objects in a natural way. Poyade et al. [10] implemented a friendly enduser Sensable Technologies PHANToM haptic interface and its usability was studied through a reassembly task in crusher maintenance. They concluded that haptic devices are a powerful interaction tool in virtual maintenance tasks even when participants felt that the haptic force feedback was far from realistic force models. The work of Gavish et al. [1] describes some design guidelines for the development of training systems for industrial maintenance and assembly tasks based on VR and/or augmented reality technologies. These guidelines were tested in some pilot tests with satisfactory results, but they were focused on the implementation of the learning strategy and not on the interaction technology used.

Recent work on virtual assembly training comes from Xia et al. [11]. They designed a multimodal system to provide trainees with more realistic mobility, as in current VEs users are constrained by a fixed position or a limited space. Thus, they designed and integrated a low cost motion simulator of human walking where Flock of Birds trackers are connected to users' feet in order to capture position and orientation. Additionally, the multimodal system is equipped with a haptic device (PHANToM Premium) as the interaction tool within the virtual assembly scene, 3D stereoscopic glasses, a spherical screen to project the virtual assembly scene, and a CyberGlove® for tracking users' hand movements. However, Xia et al. did not support the idea of combining a Data Glove with the haptic device as the Data Glove adversely affected the interaction with the VE. Instead, they proposed using a Cyberforce, which is a combination of a Data Glove with force feedback. The CyberForce may provide a realistic interaction with the VE but its main drawback is its high price. Another example of virtual assembly comes from Lu et al. [12]. For more contributions to virtual assembly training, the survey by Gupta et al. [13] is useful for works up to 2007.

Looking at haptic interaction within virtual assembly training in greater depth shows that there are few studies which analyze its benefits. Leino et al. [14] developed and implemented a haptic interface for maintenance task planning. They assessed the usability of the haptic interface in a fictitious car maintenance task where the haptic device connected Virtools simulation software with a Sensable PHANToM device. Participants reported that the haptic interface was easier to use than a 3D mouse and it improved the 6DoF navigation in a VE. However, the experiment was not analyzed further and they did not present any graphics or percentages with users' results in the usability test.

Vo et al. [15] investigated how haptic feedback affects users' performance in relevant virtual assembly tasks related to weight discrimination, part positioning, and manual assembly. Forty participants completed each task, and quantitative data was recorded. They used two PHANToM Omni configurations with two possible treatments: either with or without haptic forces (users still received visual cues in both treatments). According to the experiment's results, weight discrimination completion times were shorter with force feedback. Haptic forces also allowed greater placement accuracy when positioning virtual objects and it enabled steadier motions.

As for Bloomfield et al. [16], they researched users' performance and preferences for haptic and nonhaptic feedback for a disassembly task. There were three interactions available based on CyberGlove, PHANToM, and Spacemouse. Analysis showed that participants with the PHANToM device learned to accomplish the simulated actions significantly faster than users using the other two interaction devices. Users with SpaceMouse and CyberGlove showed more difficulty moving to specific locations. The authors did not find any significant difference on a qualitative questionnaire where users had to rate the ease with which they performed the action with the particular device.

These previous studies were focused mainly on analyzing the usability of haptic devices for assembly tasks, and few papers have explicitly analyzed the devices' effectiveness for learning tasks, i.e., the knowledge transfer achieved with these systems. Adams et al. [17] investigated the benefits of force feedback for virtual training with respect to nonfeedback technology. The simulated task consisted of assembling a LEGO biplane model composed of 29 pieces. They took an experiment with 15 people organized into three groups: (1) virtual training with haptic feedback, (2) virtual training with a 2D mouse, and (3) no virtual training. Initially, all participants watched a video tutorial of the assembly procedure. Results demonstrated that there was a learning effect in using the virtual training for those who only watched the video. Results also showed that the haptic group was significantly better than the no virtual training group the first time they performed the real task after training. Finally, the group that did not use a haptic device was not significantly different from other two groups.

A research conducted later by Oren et al. [18] studied the effect of training with a virtual system compared to a real environment. The interaction of the virtual training group used a haptic device and a data glove (one for each hand), with virtual stereoscopic visualization. The experimental task consisted of assembling a 3D wooden burr puzzle in six steps. Results demonstrated that the virtual training did not show an improvement over the real environment training.

Gavish et al. [19] analyzed the efficiency and effectiveness of different virtual training modalities based on VR technologies in a real assembly task. There were four training groups based on VR training, augmented reality training, a traditional training based on watching a video tutorial, and a traditional training based on working directly with the real assembly. Forty technicians took part in the experiment, ten in each, and were completely randomly assigned to one of the training groups. Results demonstrated that training in the VEs took more time than traditional training, but they did not find significant differences in the performance of the real task which was accomplish after the trainings. They also discussed that the use of new interaction paradigms need time to learn and it could have affected the training efficiency.

In summary, previous research suggests that the main benefit of haptic interaction versus nonhaptic interaction is shorter times for the virtual task completion. There is no evidence that having force feedback improves the learning process for performing the real task. Furthermore, the nonhaptic feedback analyzed has been limited either to 3D mouse or to haptic devices without force feedback. Hence, our research extends previous contributions by: (1) analyzing the haptic interaction versus a novel and low-cost interaction based on a Mmocap system, and a traditional interaction based on a 2D mouse; and (2) evaluating not only the usability but also the effectiveness of these interaction technologies for the learning of the real task. Exactly, we study how the learning process is affected by factors as 2D versus 3D interaction, haptic versus nonhaptic feedback, active versus passive interaction, or novel versus classical interaction technology.

## Experiment Design

This section describes the experiment conducted to study how the interaction technology affects the learning of an assembly procedural task, detailing the experimental setup, experimental task, experimental design, procedure, and data gathered.

**Experimental Setup.** The experiment was conducted on a previously published demonstrator known as IMA [20]. IMA is a controlled multimodal training system for learning assembly and disassembly procedural tasks. It supports the approach of learning by doing by means of active multimodal interaction with the virtual objects, i.e., trainees can interact and manipulate the components of the virtual scenario and simulate assembly and disassembly operations.

*Virtual System.* The IMA platform consists of a screen displaying the 3D graphical scene corresponding to the maintenance task, the device used for the interaction with the virtual scene, and the training software to simulate and teach assembly and disassembly tasks, see Fig. 1, left.

The 3D-graphical scene is divided into two areas. The virtual machine is rendered in the center of the scene and the pieces that will be assembled are placed at the back-wall. On the right edge of the screen there is a configurable "tools menu" with the virtual tools that can be chosen to accomplish the different task operations. Throughout the training session, the system provides different types of information about the task, such as: "task progress," technical descriptions of the components and tools, critical information about the operations and detailed description of errors. Critical information is also sent through audio messages. The system also automatically logs information about the task execution in order to evaluate the trainees' performance and evolution.

The platform provides guidance based on indirect or direct hints that help the users during their training process. This guidance is provided only when requested by the trainees. The indirect hints (see Fig. 1, middle) provide information about the current subtask by means of (1) visual hints: showing a second window on the screen with a copy of all the pieces involved in the current subtask in their final position and (2) textual messages: displaying both the name and description of the current subtask/step and a dynamic list with the names of the pieces/tools involved in the current step. If this aid is activated and the trainees do not yet know how to continue with the task, they can request an additional aid that provides direct information (see Fig. 1, right) about the immediate action that they should perform. For example, in the right-hand image in Fig. 1 the target piece/tool is highlighted with a different color, and in case of having a haptic device, the trainees receive an attraction force to the target object position.

*Interaction Technologies.* The IMA platform allows the trainees to interact with the VE by means of different technologies such as haptic devices of any kind, a 2D mouse, and/or Mocap based systems.

The Mmocap implementation is based on Unzueta's work [21], where the user's body part is tracked by applying the Condensation algorithm [22] with a blob representation [23]. In order to improve its tracking recognition, the users wore a chrome-key glove. A Kalman filter was efficiently applied against white noise. The system was implemented to work with one or more standard cameras, and with a stereo camera. The output is in 2D or 3D, depending on the number of cameras.

The technologies used in this experiment were based on a standard 2D mouse, a floor-grounded haptic device known as

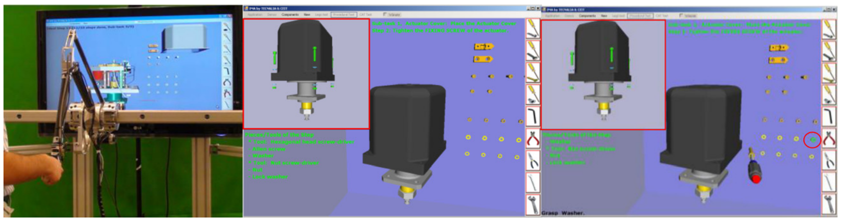

**Fig. 1 The IMA system. Left: using the LHIFAM haptic device to interact with the virtual objects and perform the assembly task. Middle: indirect aids provide information about the current subtask. Right: direct aids (in addition to indirect aids) provide information about the next action.**

LHIFAM (see Fig. 1, left), a Markerless Motion Capturing system with 2D tracking (Mmocap2D) and a Markerless Motion Capturing system with 3D tracking (Mmocap3D).

*Commands.* This experiment follows the Wizard of Oz strategy, i.e., the user does not know that the system is partially operated by a person. In this strategy, when the trainees want to send an order to the system (e.g., grasp a piece) they send the order by voice and the evaluator activates the corresponding command in the system.

**Experimental Task.** The selected experimental task consisted of assembling part of an electrohydraulic valve. This task consisted of 23 steps grouped into 5 subtasks (see Fig. 2): place the two level sensors, fix the support plate, place the electronic board, place the actuator cover and finally place the clamp. Some examples of the operations participants had to perform are: to place an electronic card in a support plate by hand, tighten screws with nuts and a wrench, plug in cables, fix cables with a cable tie, etc.

In total, participants had to manipulate 48 pieces, some of which were very delicate, and use five different tools.

In this context, an operation is classified as unimanual or bimanual depending on the number of tools needed at the same time. Thus a unimanual operation involves just one piece or tool at every step of the assembly process, even if a second hand is needed for support, while a bimanual operation requires manipulating two pieces/tools simultaneously, i.e., two hands working simultaneously.

According to the single channel theory in the motor control literature [24], humans can only process one stimulus–response at a time. In the case where humans try to do two tasks simultaneously, one task will be blocked and treated as interference. Further experiments have revealed that parallel processing on humans can occur during the early stages of processing, though single channel bottleneck occurs while selecting and programing responses. Thus, due to the way humans mentally process responses [24] and because the focus of this application was not to transfer psychomotor ability but to transfer procedural knowledge, in this study the bimanual operations were performed sequentially. First, the user chooses the proper tool or piece and places it in the final position, and then takes the second tool or piece and does the same.

**Experimental Design.** The experiment follows a between-subjects design with 60 volunteers randomly divided into five experimental groups. In four groups participants were trained on the task using the IMA VR platform, where each group used a different interaction technology: haptic interaction, mouse interaction, Mmocap2D interaction, or Mmocap3D interaction. The fifth group was trained with a demonstration video that showed how to perform each step. Unfortunately, three participants could not attend the second day of the experiment (the real task execution), so their data were not considered for the analysis. In the end there were three groups with data from 11 participants (mouse, haptic, and video) and another two groups with data from 12 participants (Mmocap2D and Mmocap3D). Participants were male and female and ranged between 18 and 60 years old, and had no prior experience with using haptic/Mocap systems. They could all perform basic assembly operations which was necessary for the experimental task. The majority of the participants had some kind of engineering degree (i.e., computer or industrial), and a few were from a different field (such as environmental sciences, physics, or secretarial studies).

**Procedure.** The experiment was undertaken on two consecutive days to avoid the short-term memory effect; on the first day the training session was held and on the second day the real task was performed. All groups followed the same protocol. First, the evaluator introduced the purpose of the evaluation. Second, participants in the virtual training groups attended a familiarization session with the IMA VR platform and the corresponding interaction technology. This session included a practical and guided explanation, with a simple example, about the features of the system (e.g., the screen layout, how to grasp/manipulate pieces/tools). Finally, participants were trained on the target assembly task. After a general explanation of the task, including the visualization of a picture of every subtask in its final state, each participant was reminded that the learning goal was to be able to perform the real task the day after. Then, all the groups had two training trials. The training trials consisted of assembling virtually the valve using the IMA VR platform, and having direct and indirect hints as support when they needed. The evaluator was in the room together with the trainee and recorded all his/her comments and problems. Information about the trainee's performance and actions was logged automatically by the VR platform.

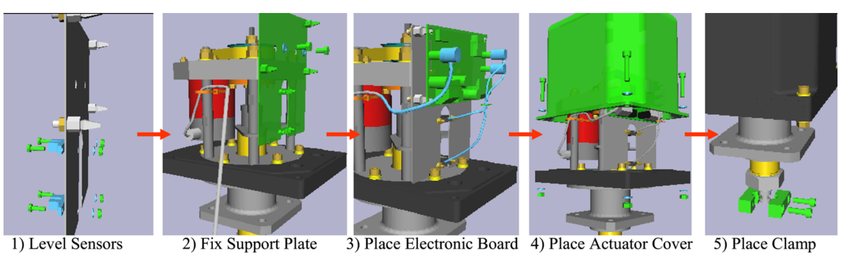


**Fig. 2 The five subtasks of the experimental assembly task**

On the second day, participants assembled the real actuator valve. They were asked to perform the assembly task correctly on their own. If they did not know what to do they could ask for evaluator support and this action was considered as an aid. The evaluator filled in a score-sheet with the relevant measures (times, errors, aids, etc.) and trainee problems that could be useful for analyzing and interpreting the results. Finally, once the participants completed the experiment, they were asked to fill out a demographic questionnaire and a usability questionnaire in order to rate the experience with the VE and the interaction technology.

**Performance Measure.** In order to analyze the participants' performance, relevant information was recorded during the different phases of the experiment. For each one of the training trials, the VR system recorded the following information: total time, overall performance (summary of the step performance including number of steps with/without errors/aids, number of total errors and number of direct/indirect aids) and detailed performance for each step (step time, type of performance—with/without errors/aids—number of errors and number of direct/indirect aids). During execution of the real task (Fig. 3), the evaluator recorded the performance time, the performance for each step (correct step without any aid/error, correct step with any solved error, correct step with any aid or incorrect step with an unsolved error), the number of nonsolved/solved errors and their type, and the number of aids.

Finally, the usability questionnaire collected the participants' feedback about the interaction technology in terms of:

Q1-Naturalness of interaction
Q2-Ease of moving/manipulating objects
Q3-User's concentration on the task
Q4-Level of difficulty of the interaction
Q5-Comfort
Q6-Ability to manage virtual objects
Q7-Consistency between the virtual and real experience

Additionally, there were two questions about the training aspects:

Q8-What percentage of the task do you think you have learnt?
Q9-Overall rating of the platform as a training system for procedural tasks

## Results

Results are presented in five subsections relating to virtual training times, real task completion times, learning of the real task, transition from the virtual task to the real task, and usability of the virtual training system. All analyses were run with the aid of minitab.

**Training Times.** In order to evaluate a training system it is important to consider not only the knowledge acquired but also the time spent using the training system. Hence, we analyze the training times in order to evaluate the efficiency of each interaction technology for learning assembly tasks.

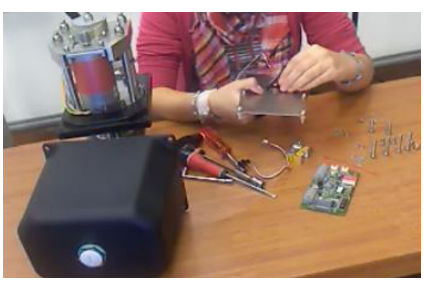

**Fig. 3 The post-training test: participants had to assemble the real valve on their own**

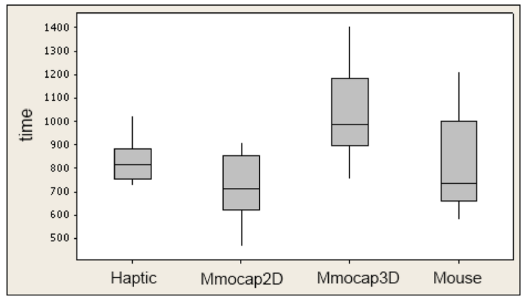


**Fig. 4 Boxplot of training times in session 1 (in seconds)**

In the following subsections, the results are presented for each training sessions.

*Training Session 1*. Participants had to perform the virtual assembly ask, knowing they could request indirect/direct aids whenever they needed them.

The decision whether and when to request guidance was completely up to the users. Figure 4 depicts the completion times for the virtual assembly training (session 1) under each interaction condition. The training session with the traditional group took 600 s, which corresponded to the duration of the demonstration video.

A one-way between-subjects analysis of variance (ANOVA), adjusted for ties, was run to compare the effect of the interaction technology on the training time. The interaction technologies analyzed were haptic, Mmocap2D, Mmocap3D, and mouse. The main effect that group had on the dependent variable, training time, during first training session was found to be significant ($F(3, 42) = 7.76$, $p = 0.000$). Tukey post hoc comparisons of the four groups indicate that Mmocap3D ($M = 964.2$, $SD = 180.6$) was significantly slower than Mmocap2D ($M = 674.2$, $SD = 135.7$), haptics ($M = 759.8$, $SD = 78.6$), and mouse ($M = 767.8$, $SD = 187.6$) conditions.

In the case of users trained under traditional training ($M = 600.0$, $SD = 0$), participants took (overall) less time than virtual training.

*Training Session 2*. Figure 5 depicts the training completion times for the virtual assembly training for each interaction condition. The training session with the traditional group took 600 s, which corresponded to duration of the demonstration video.

A one-way between-subjects ANOVA, adjusted for ties, was also run to compare the effect of the interaction technology on the training time. The main effect that interaction technology had on the dependent variable (training time) was found to be significant ($F(3, 42) = 9.39$, $p = 0.000$). Tukey post hoc comparisons of the four groups indicate that the Mmocap3D ($M = 599.3$, $SD = 118.7$) and haptics ($M = 579.8$, $SD = 90.5$) were significantly slower than the Mmocap2D ($M = 422.6$, $SD = 91.5$), and mouse ($M = 438.2$, $SD = 107.1$) groups, $p < 0.05$. Users trained under traditional training ($M = 600.0$, $SD = 0$) needed more time than virtual training under Mmocap2D and mouse conditions, and it was similar to Mmocap3D and haptic conditions.

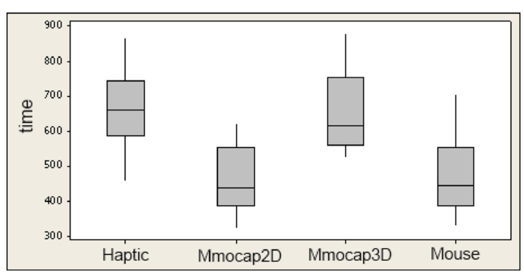


**Fig. 5 Boxplot of training times in the session 2 (in seconds)**

*Real Task Completion Time*. On the second day of the experiment, participants had to perform the real task (the same as the task learned in the virtual platform or video) and assemble the real valve with the knowledge they acquired during the training sessions. Figure 6 shows a box plot of the mean times taken to assemble the real valve. Nevertheless, it is important to take into account that participants were told to be more concerned with assembling it correctly than the time needed to assemble it.

A one-way between-subjects ANOVA, adjusted for ties, was run to compare the effect of the training condition on the real task completion time. The training conditions analyzed corresponded to haptic ($M = 858.4$, $SD = 271.0$), Mmocap2D ($M = 871.1$, $SD = 216.6$), Mmocap3D ($M = 971.1$, $SD = 207.6$), mouse ($M = 925.5$, $SD = 280.6$), and traditional training ($M = 778.6$, $SD = 265.2$).

There was no significant differences in real valve completion times among the five training conditions ($F(4, 52) = 0.97$, $p = 0.429$).

*Learning of the Assembly Procedural Task*. The learning of the assembly procedural task (Fig. 7) was measured by the percentage of the task done correctly and without any aid from the evaluator, which means via the indicator defined by

$$\text{Performance indicator} = \frac{\text{Number of correct steps (without nonsolved errors) and without any aid}}{\text{Total number of steps}}$$

A Kruskal–Wallis test was conducted to evaluate differences in the performance indicators among the four interaction types (Mmocap3D [$M = 78.44\%$, $SD = 13.5$], Mmocap2D [$M = 75.69\%$, $SD = 12.03$], mouse [$M = 83.33\%$, $SD = 11.11$], and haptic [$M = 75.76\%$, $SD = 10.18$]) in the virtual training conditions, as well as the traditional training condition ($M = 77.27\%$, $SD = 15.41$). The test, which was adjusted for ties, was found to be not significant $\chi^2(4, N = 57) = 2.56$, $p = 0.634$.

Analyzing the errors made by the participants, we defined three categories or types of error:

- Forget one step: for example, forget to tie the nylon cable
- Wrong position (or placement) of components: wrongly oriented sensors, wrong position of cable connectors (up and down), wrong position of the support, etc.
- Wrong attachment of components: forget a washer/lock washer, switch the positions of the washers and lock washers in the cover, place the washer together with the screw instead of with the nut, etc.

A step that was performed wrongly was classified into one error type based on the nature of the step. Figure 8 shows the type of errors made in each group. The analysis performed shows that the typology of the errors made in each group was similar. Specifically, the step with more errors in all groups was the placement of the sensors.

*Transition From the Virtual Task to the Real Task*. One of the greatest potential risks of VR training systems is that sometimes trainees become increasingly dependent on certain features of these systems (e.g., the use of extra cues) or on the technology interaction (e.g., simplification in the manipulation of the virtual objects), which may inhibit participants' ability to perform the real task when these features are no longer available. In order to analyze this possibility, we studied the change between the virtual task and the real task performances.

This evolution was measured as the difference in the performance indicators (i.e., the number of steps unknown) for the real and virtual task, meaning that evolution = performance indicator (real task) − performance indicator (virtual task).

Figure 9 summarizes the change in performance for all users. The results for haptic and Mmocap2D conditions showed a higher number of participants with a positive transfer of knowledge from the virtual training to the real task performance. Specifically, the 82% of users trained under the haptic condition and the 75% of users trained under the Mmocap2D condition performed better in the real task than in the virtual task. In case of the mouse and Mmocap3D conditions only the 54% and 50% of participants, respectively, demonstrated a positive transfer of knowledge from the virtual training to the real task performance.

Further analysis of these two groups (mouse and Mmocap3D) demonstrated that only 18% of participants trained with the mouse group had a negative evolution, which is similar to haptic group (18%) and Mmocap2D (16%) and significantly better than Mmocap3D group (41%).

*Usability Questionnaire*. The usability questionnaire followed a 7-Likert Scale for all the questions except for questions 7 and 8, which were rated from 1 to 10. The results are presented in Fig. 10.

A Likert scale item is a set of ordered categories, which we analyzed by means of nonparametric tests. Thus, several Kruskal–Wallis tests were run, with adjusted ties, to evaluate differences among the four interaction types (Mmocap3D, Mmocap2D, mouse, and haptic) on median change in users' experience (i.e., usability). For Q6, ability to manage virtual objects, the test was found to be significant, $\chi^2(4, N = 57) = 8.00$, $p = 0.087$. Tukey

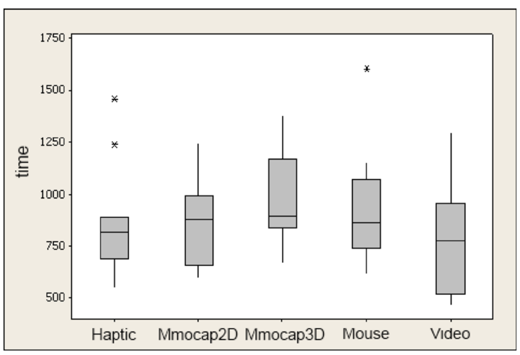


**Fig. 6 Boxplot of the times to assemble the real valve in seconds**

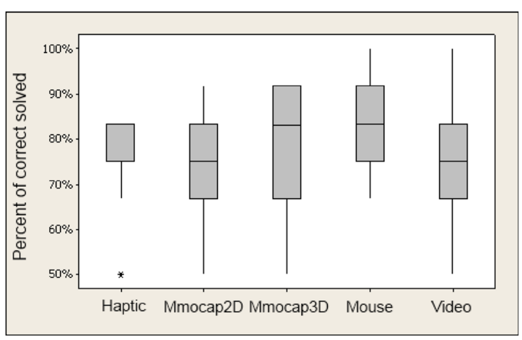


**Fig. 7 Boxplot of the performance indicators (% of correct task) for each training condition**

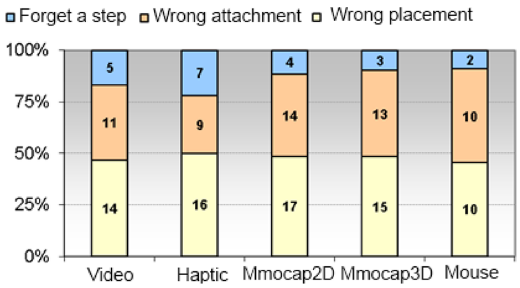


**Fig. 8 Typology of errors made in each group. Numbers inside the vertical bars represent the number of errors done in each category.**

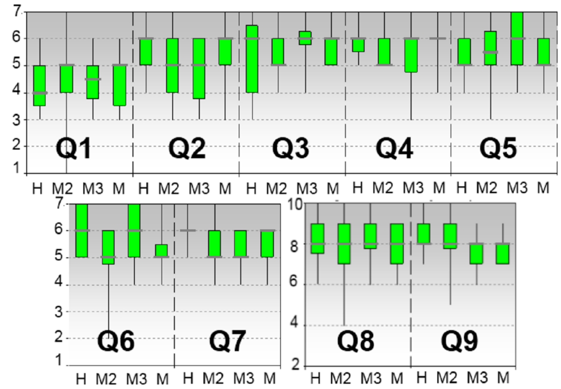


**Fig. 10 Median and quartiles values for each usability question. Box corresponds to 95% confidence interval. Q1–Q9 correspond to the usability questionnaire described in Experiment Design section.**

post hoc comparisons of the four interaction conditions did not find significance differences. For Q7, consistency between the virtual and the real environments, the test was found to be significant, $\chi^2(4, N=57)=8.00$, $p=0.046$. Tukey post hoc comparisons of the four interactions conditions found the haptic interaction to be significantly more consistent in users' opinion than both Mocap based interactions. Tests for the rest of the usability terms were not found to be significantly different among the four conditions.

## Discussion

This section presents the five subsections relating to virtual training times, real task completion times, learning of the real task, transition from the virtual task to the real task, and usability of the virtual training system.

**Virtual Training Times.** Participants performed and finished the virtual task without major problems. The virtual training based on Mmocap2D and mouse interactions were shown to be significantly faster than the Mmocap3D group during the first and second training sessions.

In analyzing the fastest groups, what they had in common was that interaction with the virtual scene was done in 2D (although the virtual scene was in 3D). To interact in the 3D space involves greater complexity, especially because it requires users to deal with depth. Moreover, knowing the depth of the user's hand or the virtual object is not so intuitive. This experiment was carried out with a standard visualization since previously we had tried low cost solutions for stereo visualization, such as anaglyphs, and we did not perceive any significant advantage. To help the users determine objects' depth, the shadows corresponding to the user's hand and the virtual objects were displayed on the floor.

As for the haptic group, the two training sessions provided different results. In the first session, the haptic training time was similar to the mouse and the Mmocap2D groups, and it was faster than the Mmocap3D. However, in the second session it was similar to the Mmocap3D, and significantly slower than the Mmocap2D and mouse. Since the Mmocap3D and haptic groups are interacting in the 3D space and have similar movement amplitudes, we should expect similar results in both training sessions.

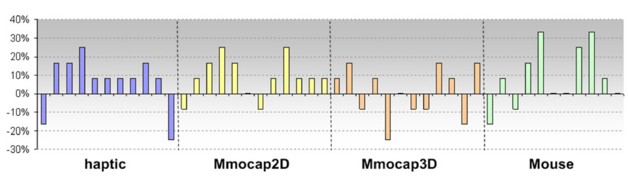


**Fig. 9 Depiction of performance evolution from the virtual task to the real task for each interaction condition. Results are presented as percentage of correct steps without any aid or error for each participant (vertical bar).**

The main reason for this difference is due to the attraction forces that were available when the user asked for direct aids in the haptic group. These attraction forces take the user's hand directly to the target piece/tool/position. Thus, the more direct hints requested, the faster the task was completed. In the experiment, the number of direct aids in the first session, when the users did not know the task, was much higher than in the second session, where users only asked for aids when they were stuck on a step.

This reduction in the number of direct hints may have affected performance times. Given these results, we can conclude that under similar conditions Mmocap3D and haptic-based interactions take similar completion times.

**Real Time Performance.** There were no significant differences in the real task performance time among groups. But this result is not conclusive since it could have been influenced by the instructions given to the participants, i.e., it was more important to assemble correctly than assemble quickly.

**Learning of the Assembly Procedural Task.** There were no significant differences in the learning acquired among the groups (i.e., traditional training, and virtual training under Mmocap2D, Mmocap3D, haptic, and mouse conditions). Nevertheless, it is important to take into account that the mouse interaction was the only interaction technology that was known and frequently used by the users. This factor could have favored their concentration on the procedural task and thus facilitated their learning.

As for the traditional training, we took it as reference. Results showed that, compared to the virtual training, learning was similar. But a few aspects should be considered. First, the video tutorial used was very clear and showed very explicitly how to perform each step of the task, and the tools and pieces that appear in the video were the same ones used in the real task.

However, in the virtual task the representation of some pieces and tools were not exactly the same as in the real task, as there was not a direct mapping between the pieces/tools in the virtual and real tasks. Second, VR systems, such as the IMA platform used in this experiment, and the tested technologies, except for the mouse, provide new interaction paradigms that can be initially complex, and users need sufficient time to learn to use them efficiently. The results of the virtual training groups could have been better if the participants had had more practice and experience using the platform and technologies. All those factors surely had an effect on the real performance, so at this stage we cannot reject the possibility that the virtual training had added value over the traditional training. Further evaluations should be performed.

Finally, we would like to come back to the research by Adams et al. [17], who studied the benefit of haptic technology versus nonhaptic technology. The strategy they followed was different from ours in that users first underwent nonvirtual training (video tutorial). Moreover, the procedural task was not an industrial assembly task (it was a Lego model assembly task) so it did not involve interacting with tools, and the necessary declarative knowledge was more limited. Initially, we thought that these changes in strategy could have led to different conclusions. However, our results are similar to theirs in that the use of haptic forces has not improved the learning of a procedural task.

**Transition From the Virtual Task to the Real One.** The main objective of virtual training systems is not to perform the virtual task correctly but to allow the knowledge needed to perform the real task correctly to be transferred. Thus, it is important to keep users from becoming dependent on the virtual training system, which could impede the knowledge to be transferred to the real case. The majority of the users showed a positive change (Fig. 9), even though the real task was performed one day after the virtual training. Since there is a temporal restriction, users could have forgotten part of the virtual training. The results for haptic, Mmocap2D, and mouse groups showed a positive transfer of knowledge from the virtual training to the real task performance in the majority of participants. However, results for the Mmocap3D group showed a positive transfer of knowledge from virtual to real task performance only in 50% of participants, while in 41% of them had a negative transfer.

**Usability of the Virtual Training Systems.** The results of the usability questionnaire were quite satisfactory. Users rated the four available interactions similarly in terms of naturalness, comfort, and easiness. The results relating to comfort are doubtful, since some users expressed symptoms of tiredness under the Mmocap3D, Mmocap2D, and haptic conditions. Maybe this tiredness problem was partly hidden due to the duration of the training session, which did not last more than 15–30 min per user. But for longer training tiredness should be studied in greater detail. If we compare the previous conditions to the mouse interaction, the movement amplitude involved is much smaller, the users can lay their hand on the table, and they can sit down. In this sense, a haptic desktop could be more appropriate for this kind of task.

Regarding the consistency of the interactions, users rated the haptic device was more highly than both Mmocap systems (2D and 3D). This can be explained from the point of view of collisions and precision. On the one hand, when there is a collision, users cannot pass through the object when using the haptic device, while with the other systems they can. On the other hand, in general tracking user movements with a haptic device allows a high degree of accuracy, on the level of millimeters, while the Mmocap based interactions lack such precision due to errors based on tracking and image processing. It is interesting that the lack of robustness in Mmocap based interactions do not seem to have affected other usability factors.

In terms of how good the virtual learning system was, the results were quite good and not affected by the interaction system; the mean rating was approximately 8 out of 10. Overall concentration on the task (Q3) and how much users learned from it (Q8) were similar independent of the interaction type and also received high ratings. Users felt that they had learnt on average around the 80% of the task. It is interesting to point out that this is precisely the percentage of the real task that was performed correctly.

These results are in concordance with the quantitative measures, where we could not find any significant difference in learning for each interaction.

## Conclusions and Future Work

We have analyzed the influence of four different interaction technologies (based on mouse, haptic device, Mmocap system with 2D tracking [Mmocap2D], and Mmocap system with 3D tracking [Mmocap3D]) in the learning of a procedural assembly task. A fifth group based on traditional training (through viewing of a demonstration video) was taken as a benchmark. In this paper we addressed the following research questions:

- RQ: How does the interaction technology affect the learning of a procedural assembly task?
  In our study, participants trained on the VR platform performed quite well (around 80% of the real task was performed correctly). The results show that there were no significant differences in the learning acquired among the four interactions nor there were significant differences with the traditional training. Therefore, it seems that the interaction technologies tested have a small or negligible impact on the learning of assembly tasks when the focus is on the transfer of procedural knowledge rather than the transfer of psychomotor skills. Two direct implications of this result are that for the transfer of procedural knowledge in assembly tasks: (1) the addition of force feedback does not increase the knowledge transfer, (2) the use of novel and low-cost Mocap systems can be a valid approach.
- RQ: Which interaction technology is more efficient, i.e., less time consuming, for a virtual training?
  Participants performed and finished the virtual task without major problems. The results show that users trained with the mouse and the Mmocap2D took significantly less training time than the users trained with the haptic and the Mmocap3D. The interaction technology can significantly affect the virtual training time, mainly due to the complexity of using it, the user's experience with the corresponding device, and the features of the workspace, e.g., small versus large movement amplitude or 2D versus 3D interaction. We found, for example, that large movement amplitude led to longer execution times than short amplitude movements did. Moreover, working directly in a three dimensional space increases complexity with respect to two-dimensional spaces, virtual navigation, and manipulation of virtual objects. In the case of haptic devices, the use of attraction forces to guide the user to the target point can decrease performance time, but the use of these haptic aids must be controlled to keep users from becoming dependent on the VE, as dependence on the VE impedes knowledge transfer.
- RQ: Does the interaction technology affect the usability of the training system?
  In general, there were no significant differences in the usability results between the interaction technologies, except for the consistency of the VE respect to the RE. This is probably a consequence of the lack of precision in Mmocap interactions. Moreover, although users did not negatively evaluate the Mmocap interactions and the haptic device with respect to comfort, some users reported tiredness symptoms during the training session. This was due to the large movement amplitude that those technologies require and the fact that users had to stand when using the technologies. A desktop haptic device could probably avoid this problem.
  From previous results we can conclude that the mouse-based interaction was more efficient for learning since the virtual training took less time, it was one of the most usable, and on average users learned slightly more than with the rest of the interactions. It is also a more popular device, it has a lower cost than the other interactions, and users did not experience the tiredness they did with the other interactions. Thus, we consider the mouse to be the best interaction for the learning of assembly tasks when the focus is on the transfer of procedural knowledge rather than the transfer of motor skills. Future work should evaluate the influence of these interaction technologies in the learning of assembly tasks when the motor component is relevant. For example, when a component requires a complex motion during assembly or when users have accessibility problems and the motion must be

precise to avoid damage to the components. Another research area could be to study these interaction technologies for the simulation of bimanual operations, where users need to work with both hands simultaneously, and therefore the use of the mouse is less natural than the other interactions. Hence, results for the other interactions could be taken into consideration, since they were designed for unimanual and bimanual task simulations.


## Acknowledgment

The authors would like to acknowledge the SKILLS project in the framework of European IST FP6 ICT-IP-035005-2006. The authors would like to thank Sara Casado, Industry and Transport Division, C/Geldo - Parque Tecnologico de Bizkaia, Derio, for her support in programing with the IMA system.


## References


[1] Gavish, N., Gutierrez Seco, T., Webel, S., Rodríguez, J., and Tecchia, F., 2011, "Design Guidelines for the Development of Virtual Reality and Augmented Reality Training Systems for Maintenance and Assembly Tasks," BIO Web of Conferences, Montpellier, France, Dec. 15–16, Vol. 1, pp. 29–33.

[2] Yuviler-Gavish, N., Krupenia, S., and Gopher, D., 2013, "Task Analysis for Developing Maintenance and Assembly Virtual Reality Training Simulators," Ergonomics in Design: The Quarterly of Human Factors Applications, **21**(1), pp. 12–19.

[3] Pere, E., Langrana, N., Gomez, D., and Burdea, G., 1996, "Virtual Mechanical Assembly on a PC-Based System," ASME Design Engineering Technical Conferences and Computers and Information in Engineering, Irvine, CA, Aug. 18–22.

[4] Brough, J. E., Schwartz, M., Gupta, S. K., Anand, D. K., Kavetsky, R., and Petterson, R., 2007, "Towards the Development of a Virtual Environment-Based Training System for Mechanical Assembly Operations," Virtual Reality, **11**(4), pp. 189–206.

[5] Schwartz, M., Gupta, S. K., Anand, D. K., and Kavetsky, R., 2007, "Virtual Mentor: A Step Towards Proactive User Monitoring and Assistance During Virtual Environment-Based Training," Performance Metrics for Intelligent Systems (PerMIS) Workshop, Gaithersburg, MD, Aug. 28–30, pp. 280–287.

[6] Bhatti, A., Nahavandi, S., Khoo, Y. B., Creighton, D., Anticev, J., and Zhou, M., 2009, "Haptically Enable Interactive Virtual Assembly Training System Development and Evaluation," International Design Engineering Technical Conferences (SIMTECT) and Computers and Information in Engineering, Adelaide, Australia, June 15–19, pp. 1–6.

[7] Abate, A. F., Guida, M., Leoncini, P., Nappi, M., and Ricciardi, S., 2009, "A Haptic-Based Approach to Virtual Training for Aerospace Industry," J. Vis. Lang. Comput., **20**(5), pp. 318–325.

[8] Seth, A., Su, H. J., and Vance, J. M., 2005, "A Desktop Networked Haptic VR Interface for Mechanical Assembly," Proceedings of IMECE, Orlando, FL, Nov. 5–11, pp. 1–8.

[9] Seth, A., Su, H. J., Vance, J. M., and Fellow, A., 2008, "Development of a Dual-Handed Haptic Assembly System: Sharp," ASME J. Comput. Inf. Sci. Eng., **8**(4), p. 044502.

[10] Poyade, M., Reyes-Lecuona, A., Leino, S., Kiviranta, S., Viciana-Abad, R., and Lind, S., 2009, A High-Level Haptic Interface for Enhanced Interaction Within Virtools, Virtual and Mixed Reality, Third International Conference, VMR 2009, Held as Part of HCI International, San Diego, CA, July 19–24, pp. 365–374.

[11] Xia, P., Lopes, A. M., Restivo, M. T., and Yao, Y., 2012, "A New Type Haptics-Based Virtual Environment System for Assembly Training of Complex Products," Int. J. Adv. Manuf. Technol., **58**(1–4), pp. 379–396.

[12] Lu, X., Qi, Y., Zhou, T., and Yao, X., 2012, "Constraint-Based Virtual Assembly Training System for Aircraft Engine," *Advances in Computational Environment Science, Advances in Intelligent and Soft Computing*, Springer, Berlin, Vol. 142, pp. 105–112.

[13] Gupta, S. K., Anand, D. K., Brough, J. E., Kavetsky, R. A., Schwartz, M., and Thakur, A., 2008, "A Survey of the Virtual Environments-Based Assembly Training Applications," Virtual Manufacturing Workshop, UMCP, Turin, Italy.

[14] Leino, S., Lind, S., Poyade, M., Kiviranta, S., Multanen, P., Reyes-Lecuona, A., Mäkiranta, A., and Muhammad, A., 2009, "Enhanced Industrial Maintenance Work Task Planning by Using Virtual Engineering Tools and Haptic User Interfaces," Virtual and Mixed Reality, Vol. 5622 (Lecture Notes in Computer Science), Springer, Berlin, pp. 346–354.

[15] Vo, D. M., Vance, J. M., and Marasinghe, M. G., 2009, "Assessment of Haptics-Based Interaction for Assembly Tasks in Virtual Reality," Third Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Salt Lake City, UT, Mar. 18–20, pp. 494–499.

[16] Bloomfield, A., Deng, Y., Wampler, J., Rondot, P., Harth, D., McManus, M., and Badler, N. I., 2003, "A Taxonomy and Comparison of Haptic Actions for Disassembly Tasks," IEEE Virtual Reality Conference (VR 2003), Los Angeles, CA, Mar. 22–26, pp. 225–231.

[17] Adams, J. R., Clowden, D., and Hannaford, B., 2001, "Virtual Training for a Manual Assembly Task," Haptics-e, **2**(2), pp. 1–7.

[18] Oren, M., Carlson, P., Gilbert, S., and Vance, J. M., 2012, "Puzzle Assembly Training: Real World vs. Virtual Environment," IEEE Virtual Reality, Orange County, CA, Mar. 4–8, pp. 27–30.

[19] Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., and Tecchia, F., 2013, "Evaluating Virtual Reality and Augmented Reality for Industrial Maintenance and Assembly Tasks," Interact. Learn. Environ. J, pp. 1–21.

[20] Gutiérrez, T., Rodríguez, J., Vélaz, Y., Casado, S., Suescun, A., and Sánchez, E. J., 2010, "IMA-VR: A Multimodal Virtual Training System for Skills Transfer in Industrial Maintenance and Assembly Tasks," IEEE 19th International Symposium on Robot and Human Interactive Communication (RO-MAN), Pisa, Italy, Sept. 13–15, pp. 428–433.

[21] Unzueta, L., 2008, "Markerless Full-Body Human Motion Capture and Combined Motor Action Recognition for Human–Computer Interaction," Universidad de Navarra, Tecnun, San Sebastián, Spain.

[22] Isard, M., and Blake, A., 1998, "Condensation—Conditional Density Propagation for Visual Tracking," Int. J. Comput. Vision, **29**(1), pp. 5–28.

[23] Wren, C. R., Azarbayejani, A., Darrell, T., and Pentland, A. P., 1997, "Pfinder: Real-Time Tracking of the Human Body," IEEE Trans. Pattern Anal. Machine Intell., **19**(7), pp. 780–785.

[24] Schmidt, R. A., and Lee, T. D., 2005, *Motor Control and Learning: A Behavioral Emphasis*, Human Kinetics Publishers, Inc., Champaign, IL.