

Multimodal Evaluation of the Differences between Real and Virtual Assemblies

Mikel Sagardia and Thomas Hulin

Abstract—What are the technological bottlenecks in virtual assembly simulations with haptic feedback? To tackle this question, we present an evaluation study in which real feedback modalities are gradually replaced by synthetic ones. In particular, the effects of the following factors on the user performance and perception during virtual assemblies are analyzed: (i) a visual feedback system consisting of an nVisor head-mounted display, (ii) our haptic device HUG suited for unscaled upper-body movements, and (iii) our novel six-DoF constraint-based haptic rendering algorithm. Besides of that, the influence of (iv) real collision sounds is also examined to a lesser extent. The experimental assembly scenario consisted of three variations of peg-in-hole tasks which were performed by a total of $N = 24$ participants in a within-design study. The mentioned three synthetic factors (i)-(iii) gradually replaced in five degrees or steps the real feedback sources, ending up in completely virtual assembly simulations. For each of the degrees, three objective variables (completion time, collision forces, and muscular effort) and five subjective ratings (related to the perception of realism and the workload) were recorded and statistically analyzed. In order to explain subjective perception also with objective measures, reaction times of a secondary audio task performed in parallel with the assembly exercises were recorded, too. While previous works have mainly focused on differences of completion time between real and virtual manipulations, our results show how all of the mentioned twelve performance and perception indicators are influenced by each of the four varied feedback factors, building a multi-modality relationship function that maps our or similar systems and expected user responses. In general, the haptic feedback modality turned out to have the largest impact on the dependent variables, particularly the HUG interface, whereas audio cues seemed to be less significant. We quantify these previous and further qualitative statements within the domain defined by the used systems. Moreover, the relationship of our insights with related other work is discussed, and their projections are outlined.

Index Terms—haptic rendering, haptic interfaces, multi-modal virtual reality, evaluation, user studies.

1 INTRODUCTION

Virtual manipulations with haptic feedback are appealing to medical [1] or manufacturing [2] applications, amongst others, since they enable real-time training and verification simulations in a great variety of environments without the need of building the scenarios physically. In particular, force feedback has been shown to be superior to visual [3], and tactile feedback [4] when displaying collisions. Yet, virtual simulations with force feedback lead to lower performances compared to the real-world experiences (see, for instance, [5] and [6]).

That difference in performance between real and virtual manipulations arises from the limited fidelity provided by the used interfaces and the virtual simulations. The model illustrated in Fig. 1 helps understand the major elements that play an important role during the perception of contacts. The internal *perceptual hypothesis* of the world is shaped with the interpretation of the *sensations* caused by the *proximal stimulus* [7]. When it comes to the perception of the world during assembly manipulations, the most relevant stimuli can be related to the modalities of (i) *visual*, (ii) *haptic*, and (iii) *acoustic feedback*. The stimuli of these feedback modalities cause sensations that can affect each other during the process of creating the *percept* [8], [9]; hence, a holistic comprehension of contact perception requires analyzing the feedback modalities, their subsystems, and their interaction.

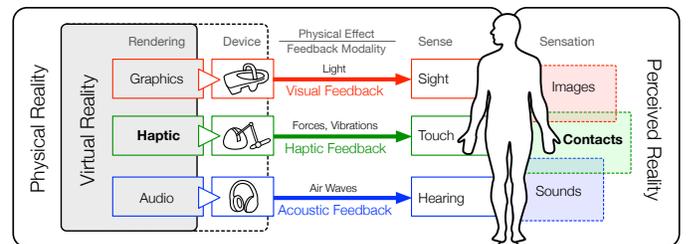


Fig. 1. Model for virtual contact perception, focusing on the senses of touch, sight, and hearing. Virtual simulations try to replicate real-world phenomena by (first) rendering and (then) displaying via devices different feedback modalities, which are sensed by human operators. The sensation of contact, related primarily to the sense of touch, can be influenced by other senses. In this study, virtual and real interactions are compared by analyzing the effects of the used rendering and device upon user performance and perception.

In general, feedback modalities can be *real* or *synthetic*, depending on whether their origin is in the *physical* or a *virtual reality* (VR), respectively. The two major underlying subsystems in feedback modalities of *synthetic* nature are (i) the *rendering* or *signal generation* and (ii) the *device* or *display medium*. Virtual simulation setups try to replicate the real physical world by means of these two constituent elements (for each feedback modality); in other words, the loss in performance and perception between real and virtual manipulations can be explained by observing the effect of these subsystems and their interplay.

Along these lines, this work tries to isolate the effect of the *synthetic visual feedback*, the *haptic device* and the *haptic*

• M. Sagardia and T. Hulin are with the Institute of Robotics and Mechatronics of the German Aerospace Center (DLR), Wessling, Germany. Contact e-mail: mikel.sagardia@dlr.de

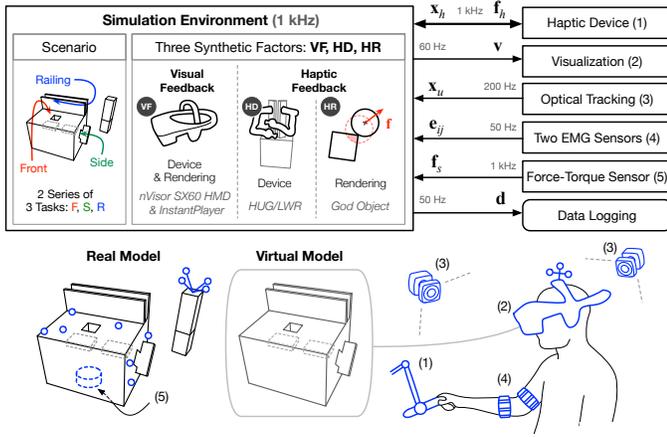


Fig. 2. General diagram of the setup in the user evaluation. The $N = 24$ participants performed with real and virtual models several exercises composed of two series of three tasks: frontal insertion, side insertion, and railing. In each of the exercises, the factors of *Visual Feedback* (VF), *Haptic Device* (HD), and *Haptic Rendering* (HR) were varied in two levels: *synthetic* and *real* (i.e., no virtual feedback provided in the *real* treatment). That variation led to five *Degrees of Virtualization* (D, see Fig. 5), from purely real to purely virtual. The order of all factor levels was systematically permuted. In the exercises where the real models were used, the trajectories were optically tracked and the contact forces registered using a JR3 force-torque sensor. Additionally, the muscular EMG signals of the participants were measured with two Myo armbands during all exercises.

rendering compared to real physical feedback on user performance and perception. The influence of *real acoustic signals* is also analyzed. For that purpose, $N = 24$ participants performed in a within-design user study three well-defined assembly tasks with real models and, gradually, the real subsystems were replaced by their synthetic equivalents, ending up with completely virtual environments. The haptic device tested in the completely virtual environment was the HUG [10] and the haptic rendering algorithm a new hybrid *god object* approach [11], also described in this work. The main contribution of this work lies on the characterization of those subsystems. In addition to comparing task completion times and subjective perception ratings, as commonly done in the literature, also force values, muscular effort, and reaction times were collated. This results in a more complete model for virtual assembly performance and perception.

This work is organized as follows: next Sec. 2 outlines some relevant evaluation studies related to this work. Then, the device and the rendering method of the used synthetic haptic feedback framework are introduced in Sec. 3. Sec. 4 describes the experimental setup and design, including the tasks performed by the participants, the varied factors, and the followed procedure. The results and discussion of the experiments are presented in Sec. 5 and, finally, Sec. 6 concludes with the most important insights.

2 RELATED WORK

Several works have been published in the past two decades in which the manipulation performance with systems that provide synthetic haptic feedback has been evaluated in comparison to real physical interaction. Gupta et al. [5] are one of the pioneers in that field. The authors evaluated

the time required for performing a peg-in-hole task in 2D both in real and virtual scenarios with haptic feedback. Two PHANToMs¹ were used attached to the thumb and index finger and the difficulty index (a combination of friction, distance, and clearance) was varied. Although the trends when varying difficulty were similar in both virtual and real environments, virtual assemblies required $\sim 2\times$ as long as real manipulations. Additionally, the authors also tested the influence of synthetic audio feedback, but it had no significant effect on the time performance.

By using their own magnetic levitation haptic device, Unger et al. [12] also tested a real and virtual peg-in-hole scenario, this time in 3D. They concluded that synthetic force feedback was worse than real feedback, but better than synthetic visual feedback alone because it provided a better sense of realism. Additionally, they observed that virtual manipulations with haptic feedback lasted $\sim 2.75\times$ longer than real ones. Bashir et al. [13] carried out a comprehensive user study in which manipulation performances using 15 synthetic modi were compared to real interactions. Among others, they varied the visualization and tested synthetic three-DoF force feedback in the experiments. The used scenario comprised also sliding tasks in addition to the peg-in-hole exercise. The authors found no significant effects due to the variation of the visualization, but they observed that virtual manipulations with force feedback were 45% slower than real ones.

In a user study conducted by Garbaya and Zaldivar-Colado [6], participants assembled concentric cylinders in real and virtual environments; for the latter modality, a CyberGrasp and a CyberForce² were used. It turned out that the virtual manipulations with haptic collision feedback led to $2\times$ longer times than in the real condition, but $2.5\times$ faster completions were achieved in comparison to virtual manipulations without haptic collision feedback. Lim et al. [14] showed that elements such as chamfers improve completion times up to 33% in peg-in-hole scenarios when performing in real and virtual scenarios with haptic feedback. In contrast, using a PHANToM Omni and stereo visualization, the performance times were roughly $\sim 4.5\times$ longer than in real environments. González-Badillo et al. [15], amongst who are also the authors of the last two cited publications, evaluated a setup composed of two PHANToM Omnis using more complex scenarios, which consisted in assemblies of a gear, a pump, a bench, and similar. The authors achieved the best completion times in the virtual environment when combining collision haptics and assembly fixture guidance; yet, the time performance was $\sim 2 - 3\times$ slower than in real environments.

As the reader can see, differences in time performance depend on the haptic devices used, and the device and the rendering are rarely evaluated separately from each other. Moreover, few works report comparisons beyond time performances. Conversely, the analysis introduced in this paper tries to give more complete insights of other relevant factors and indicators involved.

Some researchers have also provided evaluation frameworks for haptic devices. For instance, Harders et al. [16]

1. <http://www.dentsable.com/haptic-phantom-desktop.htm>
 2. <http://www.cyberglovesystems.com/>

compared three haptic devices in a virtual scenario where the users had to build a 3D puzzle. No differences in terms of task completion time were found between the three devices, which were all desktop-based and had six DoF. On the other hand, Samur et al. [17] presented and tested a series of testbeds for measuring human-machine interaction performance focusing on haptic interfaces. Recently, we have reported the results of a user study [18] in which the differences between two haptic interfaces are discussed, among other contributions. That publication is the previous work to the study presented in this paper. Using the same scenario, the haptic device HUG employed in this work was compared to the Sigma.7³. Different haptic rendering paradigms with varied stiffness were contrasted as well. The study concluded that the constraint-based approach employed in this work adjusted with a lower stiffness value than the maximum stable one yielded the best contact realism. The contact realism ratings were also increased when the HUG was used, although it led to worse performance and ergonomics values.

3 SYNTHETIC HAPTIC FEEDBACK

This section briefly presents the two subsystems of the synthetic haptic feedback used in this work: the *device* (Sec. 3.1) and the *rendering* (Sec. 3.2). The most relevant features required for better understanding the current work are provided, leaving aside deeper technical insights that can be consulted in the respectively referenced publications. Note that the haptic *rendering* method deserves especial attention; not only because it is a novel approach, but also because synthetically rendered force signals have rarely been contrasted with real contacts while removing the influence of the used mechanical interface.

3.1 Haptic Device: HUG

The haptic device used in the study is the HUG [10], shown in Fig. 3. It consists of two torque-controlled DLR/KUKA Light-Weight Robot (LWR) arms⁴ able of providing with six-DoF force feedback. Electronics and redundant sensors integrated in the seven revolute joints of each LWR are controlled at different frequencies, being 1 kHz the minimum update rate guaranteed. On the other hand, an additional force-torque sensor at the end-effector enables a feed-forward compensation that reduces up to a 33% the inertia introduced by the dynamic mass of each robot, which is $m_{LWR} = 14$ kg.

This bimanual device is characterized by its large workspace that covers the whole upper body (maximum arm span of 0.9 m) and the high forces that it can display (peak values of 150 N). In particular, the device configuration was optimized by analyzing reachability maps [19] so that its workspace maximally covers the human upper-body workspace. In the same line, thanks to the redundant kinematic of seven-DoF, the robot elbows can be optimally configured to react compliantly to external forces and optimize their position with respect to the position of the user’s elbows, avoiding singularities. All these properties make

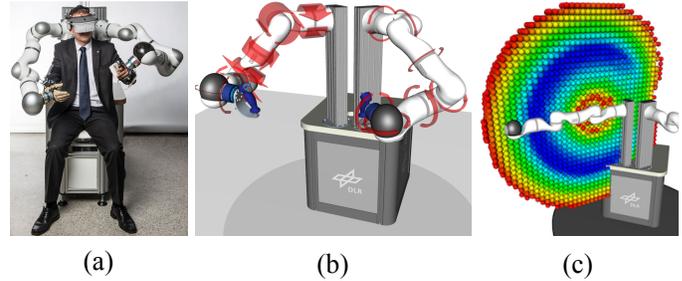


Fig. 3. (a) The HUG bimanual haptic device, composed of two DLR/KUKA Light-Weight Robots (LWR). (b) Recorded torque values in joints displayed as red rotating arrows. (c) Reachability map of the LWR; its overlap with the human reachability map was optimized for obtaining the configuration of the HUG.

of the HUG a haptic interface suited for a great variety of common and unscaled hand manipulations.

In addition to the interaction with virtual environments, the device has often been used also in teleoperation applications with humanoids [20]. Despite of the successful applications carried out with it, we are not aware of any user evaluations in which the impact of the HUG or the LWR as an input device on the user performance is analyzed. Hence, we hope this work contributes with some new insights in this regard.

It is worth mentioning that only one arm was used during the study, activated by the user with a foot-pedal. Furthermore, although several interfaces (such as data gloves, joysticks, or grippers) can be magnetically coupled to the device, users interacted with a simple handle in order to reduce additional influences and simplify the analysis.

3.2 Haptic Rendering: Hybrid God-Object Heuristic

The haptic rendering method used in this work is a *god-object* heuristic that computes six-DoF forces for arbitrarily complex rigid geometries, as presented by Sagardia and Hulin in [11]. Although it is a *constraint*-based rendering method, it requires a *penalty*-based algorithm running in the background, and it combines the advantages of both rendering paradigms: (i) real-time and robust handling of complex geometries and (ii) no or marginal interpenetration. Fig. 4 gives an overview of the data structures and the algorithm, introduced in the remaining of this section.

At the beginning of each cycle t , the penetration (p) and the penalty forces (f_p , t_p) of the constrained god object in the previous cycle ($t - 1$) are computed. The integrated penalty-based algorithm, presented by Sagardia et al. [21], is an optimized re-implementation of the well-known Voxelmap-Pointshell (VPS) algorithm [22]. Signed distances fields (*voxelmaps*) and point-sphere trees (*pointshells*) are employed during the online queries (see Fig. 4 (a,b)).

It is important to mention that so as the *god object* heuristic work correctly, the voxelmaps are slightly dilated with a value d . Whereas the penetration p gives account of the real overlap distance between the objects, points lying in the safety margin between the real surface and the isosurface dilated at d are considered to be colliding, and, therefore, used for the penalty force computation.

Fig. 4 (c) schematically illustrates the *god object* simulation that is processed upon the penalty values. The al-

3. <http://www.forcedimension.com/products/sigma-7/overview>

4. <http://www.kuka-lbr-iiwa.com>

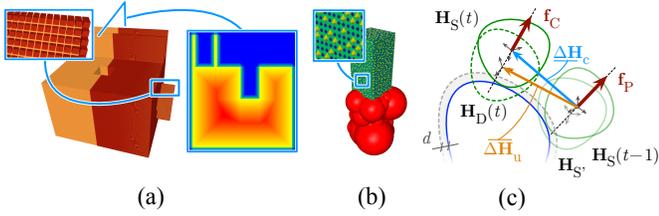


Fig. 4. Data structures and methods used for haptic rendering. (a) Partially voxelized representation of the model used in this work and its sagittal section with color-coded signed distance values (1 mm voxel edge, $253 \times 383 \times 283$ voxels). (b) Point-sphere hierarchy structure of the peg object used in this work: it consists of 5513 points distributed into 1842 clusters classified in 8 levels (two consecutive point levels and a sphere level are shown). (c) Schematic of the constraint-based haptic rendering heuristic: the movement in the direction of the contact normals or penalty forces (\mathbf{f}_P) is blocked; as a result, the object remains on the surface (\mathbf{H}_S), while the user tries to penetrate the object with the device (\mathbf{H}_D). The difference between both frames yields the six-DoF constraint force (\mathbf{f}_C , \mathbf{t}_C) displayed to the user.

gorithm consists essentially in constraining the movement of the *god object* proxy to the surface as it tries to go from its previous frame $S(t-1)$ to the current device or tool frame $D(t)$ (with homogeneous transformations \mathbf{H}_S and \mathbf{H}_D , respectively). For that, first, the residual penetration error of the previous $S'(t-1)$ is corrected, obtaining the corrected proxy frame $S(t-1)$ ($S' \rightarrow S$). Second, the unconstrained movement $\Delta\mathbf{H}_u$ of the proxy is computed. This transformation is the one which brings the proxy from $S(t-1)$ to the current device frame $D(t)$ without taking into account the contact surface. Third, the constrained movement $\Delta\mathbf{H}_c$ is built from the unconstrained allowing all perpendicular components to the six-DoF penalty forces and only the parallel components which do not point in the opposite direction to the penalty forces. In the process, Coulomb friction is also rendered by restricting the length of the constrained movement until its intersection with the desired friction cone with apex in D .

Finally, the constraint forces and torques (\mathbf{f}_C , \mathbf{t}_C) are the spring forces related to the virtual coupling (VC) transformation between the device frame $D(t)$ and the *god object* or current proxy frame $S(t)$.

The firstly computed penalty forces could be displayed to the user; however, that leads to *pop-through* and instability issues. Xu and Barbič [23] have recently tackled these problems by implementing an adaptive weighted stiffness approach. The constraint-based method outlined here yields more realistic and stable contacts than re-implemented VPS approach used alone [18], and successfully handles realistic complex geometries [24]. Additionally, the simulation runs extremely fast ($5 \mu\text{s}$) and bounds the penetration error to the used resolution.

4 EXPERIMENTAL DESIGN AND IMPLEMENTATION

Fig. 2 shows the concept of the study and illustrates the major components of the *apparatus*. The tasks carried out by the participants are outlined, as well as the *varied feedback factors* when working with the real and the virtual models. This section elaborates all those points and thoroughly describes the experimental procedure.

4.1 Apparatus and Varied Factors

The main goal of the study was to compare virtual assembly manipulations with real ones in order to identify the effect of each system component. To that end, interfaces and synthetic scenarios were introduced in different degrees, while the exercises to be performed by the subjects (Sec. 4.2) remained constant. The main focus lays on the evaluation of *haptic feedback*, but *visual feedback* was also altered in the process.

In the following, all components of the setup are described and subsequently it is discussed how the variation was implemented.

4.1.1 Setup Components

The setup basically consists of five components:

- (1) The haptic device used in the study is the HUG [10], introduced in Sec. 3.1. The used stiffness was $k = 3000 \text{ N/m}$, without added damping. These values were obtained after a previous user study [18] and during pilot tests, trying to reproduce the real contacts with the highest fidelity possible.
- (2) The visualization was powered by the Instant-Player⁵ engine from Fraunhofer IGD and displayed with an nVisor SX60⁶ from NVIS. The standard eye separation of 63 mm was adjusted for each participant for an optimum 3D vision.
- (3) A Vicon Bonita⁷ optical tracking system was used to track the movements of the real models and the head movements. On account of the tracking, the participants could move the camera view intuitively in the scenario.
- (4) Two Myo armbands⁸ from Thalmic Labs were used for recording the electromyographical (EMG) signals of the upper and forearms. Since each person has a unique pattern, we calibrated the signal values prior to the exercises with each device. To that end, the participants had to relax and leave their arm hanging (minimum reference value) and then lift a 2 kg weight with their hands for about 5 s, stretching out their arm straight with 90° between arm and chest on the frontal or coronal plane (maximum reference value). The scalar effort signal was synthesized as the two-norm of the 2×8 values streamed by the armbands. For the evaluation, the recorded effort signal was divided by the maximum calibration value after subtracting the minimum one.
- (5) The real assembly model was mounted on a JR3⁹ force-torque sensor with which physical collisions were measured.

4.1.2 Varied Factors: Degrees

Three feedback factors were defined, each one with the two levels of treatment *real* (R) and *synthetic* (S):

- VF *Visual Feedback*: the *real* treatment of this factor means the user saw with the bare eyes the real

5. <http://www.instantreality.org>

6. <http://www.nvisinc.com>

7. <https://www.vicon.com/products/camera-systems/bonita>

8. <https://www.thalmic.com>

9. <http://www.jr3.com>

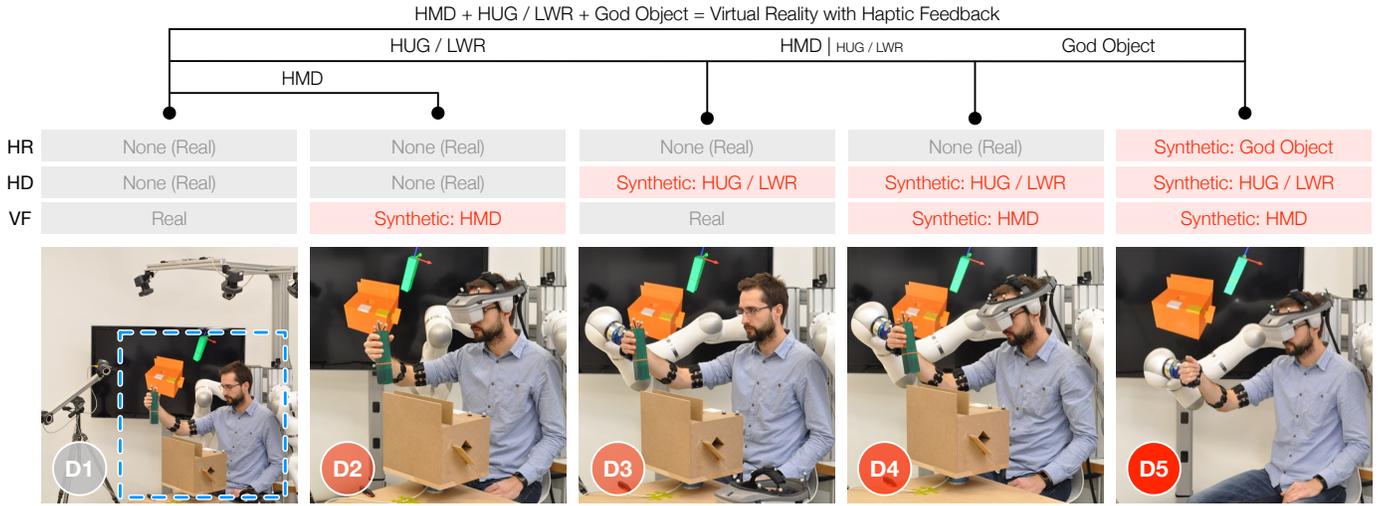


Fig. 5. Degrees of virtualization, starting from D1 (purely real) to D5 (purely virtual). The whole setup with the tracking cameras and the HUG is also displayed in D1. For each degree, the binary values (*real* or *synthetic*) of their three main factors are specified: *Visual Feedback* (VF), *Haptic Device* (HD), and *Haptic Rendering* (HR). Additionally, the comparisons between degrees are reported; e.g., the collation between D1–D2 isolates the effect of the head-mounted display (HMD).

models, whereas with the *synthetic* treatment virtual images were displayed on the HMD.

- HD *Haptic Device*: with the *real* treatment of this factor, the user had no haptic device coupled to the hand, i.e., manipulations were done with the bare hand holding the physical peg; for the *synthetic* treatment, the HUG (see Sec. 3.1) was connected to the hand of the user.
- HR *Haptic Rendering*: when the *real* treatment of this factor was applied, the users felt the collisions between the physical models; conversely, during the *synthetic* treatment, the virtual forces computed by the haptic rendering (see Sec. 3.2) algorithm were displayed to the user via the HUG. That implies that so as HR to be *synthetic*, HD needed to be, too.

As the reader can deduce, VF comprises both the *rendering* and the *device* parts of the *synthetic visual feedback* illustrated in Fig. 1, whereas the *synthetic haptic feedback* is divided into the other two factors HD (*device*) and HR (*rendering*). This allows for a better analysis of the haptic modality, as sought in the study.

As illustrated in Fig. 5, the combination of the different treatments of the three factors leads to five *degrees of virtualization*:

- D1 This *degree* can be considered as purely real. The participants performed the exercise in the physical reality without any *synthetic* feedback systems (i.e., $VF = R$, $HD = R$, $HF = R$); however, as in all other degrees, trajectories, collision forces and the muscular effort were measured (see Fig. 2).
- D2 The participants performed the exercise with the physical models but saw virtual images through the HMD (i.e., $VF = S$, $HD = R$, $HF = R$).
- D3 The participants performed the exercise with the physical models observing them with their bare eyes, but had to move the HUG arm (i.e., $VF = R$, $HD = S$, $HF = R$). The robot arm was gravity compensated and did not display any virtual forces.

Nevertheless, the users could feel the inertia of the system.

- D4 The participants performed the exercise with the physical models but moving the HUG and watching virtual images through the HMD. This *degree* is the superposition of the previous two (i.e., $VF = S$, $HD = S$, $HF = R$).
- D5 In this *degree* all factors were synthetic, thus, the participants worked with the virtual models instead of the real ones. In contrast to D1, this *degree* is purely virtual.

By comparing the dependent values obtained under each degree, the effects of the synthetic factors over the real level can be determined (see Fig. 5 above).

4.2 Tested Scenario: Tasks and Exercises

Fig. 6 shows the real and virtual models used in this study, their measurements, and the tasks performed with them. An exercise is defined to be two series of a sequence of three different peg-in-hole tasks that try to abstract common manipulation scenarios and maximize the transference of the results. The used assembly model had two reference boxes: a gray one set in the middle for starting and finishing the whole exercise (it turned blue during the exercise); and a yellow one located on the right corner for starting and finishing the tasks (it turned red during the exercise). Right after starting the exercise (collision on the middle box), the users had to carry out the 2×3 tasks in a row and without pauses. The order of the tasks was systematically permuted for each user session (and maintained constant during it) and the participants had to count aloud the tasks to ensure they did not miss any of them. Additionally, the participants were instructed to perform the tasks (*primary goal*) with the lowest contact forces possible and (*secondary*) as fast as possible. In the following, we describe the tasks and their properties (see Fig. 6):

- (i) *Frontal Insertion*: This task is the common peg-in-hole scenario; the green square peg needs to be

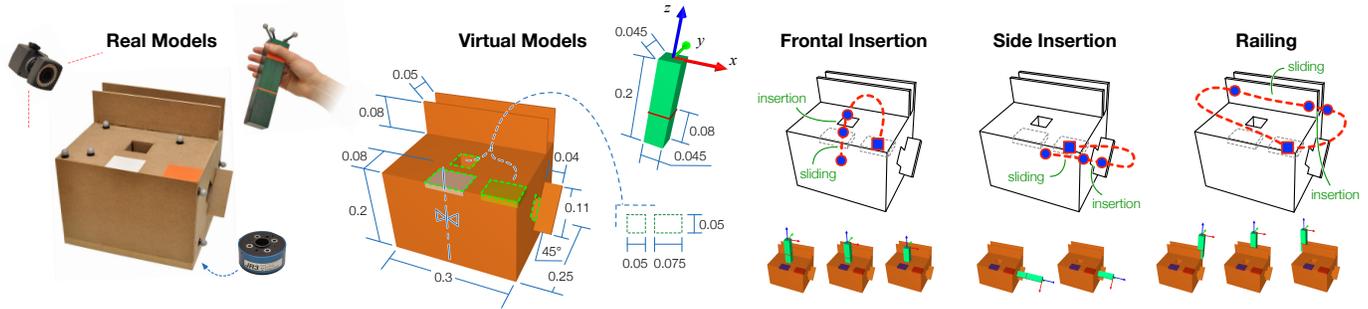


Fig. 6. Real and virtual models and tasks in the scenario of the user study (measurements in meters). Each exercise consisted of two series of a combination of three peg-in-hole tasks: frontal assembly, side assembly, and railing. The grasping regions of the peg were standardized by marking them with red stickers. All holes had a clearance of 5 mm and a depth of 8 cm, marked with a red ring on the peg. The participants had to hit a white middle-box for starting and finishing the exercise; additionally, a yellow corner-box had to be hit between the tasks. The boxes had a different color (blue and red) during each exercise in order to indicate that the process was being recorded. Pauses were not allowed between the tasks.

inserted into the upper hole, until the bottom of the hole is touched (a red ring on the peg shows the depth of the hole). Similar exercises have been long conducted by many other authors [5].

- (ii) Side Insertion: This task is similar to the previous one, but the hole is located on the right side of the model. Hence, matching the required orientation is more difficult and the visibility is not as good as in the frontal insertion.
- (iii) Railing: In this task, the green peg must be translated between two walls from the right to the left. Thus, it consists in a constrained translational task, as proposed in [25].

The three tasks require *positioning*, *insertion*, and *sliding* in different contact and manipulation configurations. Furthermore, they cover several important telemanipulation task categories identified by other authors [26], [27]. In general, all tasks involve *manipulations* with *manual dexterity* for controlling *translations* and *orientations*. In particular, the initial positioning requires *gross movements* to bring the peg from the corner to the target hole and the subsequent insertion is more related to *fine motor control*. On the other hand, both insertion and sliding subtasks require a higher *perception* or awareness of the forces that are occurring.

Therefore, as mentioned, with this scenario it is intended to comprise many features present in common manipulation tasks, while keeping the geometries as abstract as possible to derive more general and transferable findings.

4.3 Secondary Task and Auditory Privation

The effects of two more variations were tested and evaluated also: (i) the influence of a *secondary task* that had to be performed in parallel and (ii) the impact *auditory privation* during the exercises.

The first consisted in a short (~ 200 ms) and loud horn or “beep” that was systematically played in cycles during the exercises, without perceivable periodicity. The users had to press a pedal with their left foot as soon as they heard the sound and the reaction times were measured with millisecond accuracy. Each *secondary task* cycle lasted 8 s and the sound was played in a random instant during the period of 2–8 s. In addition to the regular degree exercises explained in the previous section, the participants had to carry out all

degrees with this *secondary task* too, and, during them, they were told that both assembly exercises and *secondary tasks* performed in parallel had the same priority. Reaction times to audio signals conceived as *secondary tasks* have been used in the literature as objective measures of workload [28] or presence [29].

Second, participants had to carry out the exercises wearing ear plugs and headphones with active noise-cancellation; additionally, white noise was played on the headphones in order to maximally obstruct their auditory perception. Exercises with this *auditory privation* were carried out only during the degrees with *real* visual feedback (i.e., D1 and D3), because no full obstruction with headphones was possible in practice if the users wore the HMD. Although it has been suggested that there is no specific dominance in tri-sensory (vision-audio-haptic) tasks [9], haptic stiffness perception has been shown to be biased by real [30] and synthetic [31] sound cues. Hence, it is relevant to analyze to what extent *real* audio cues (or their absence) might affect the user interaction with the presented setup.

4.4 Sample, Procedure, and Collected Data

A total of $N = 24$ participants with marginal¹⁰ or no virtual assembly experience were recruited. The statistically standard participant was male (3 female), right-handed (3 left-handed), and $M_{\text{age}} = 28.79$ years old ($Md_{\text{age}} = 27$). All participants had a university degree or (6 of them) were undergraduate students. All subjects read and signed a consent form and they were not paid for their participation.

All participants tried all conditions (within-design) in one session that roughly lasted 1.5 h (including explanations and pauses). The procedure followed with each participant can be summarized in the following steps:

- #1 Standardized **instructions** were given to the participants in form of slides. The purpose of the study was explained, the procedure, and a video of the exercise. Additionally, the participants learned how to use the devices (~ 15 min).
- #2 The **consent form** and a demographic and experience questionnaire were filled out by the participants (~ 5 min).

¹⁰ Eight of them participated in a maximum of three previous user studies with similar systems in the past three years.

- #3 The participants carried out 2 **test trials** with the real models until they felt comfortable with the order of the tasks of the exercise (~ 2 min).
- #4 For each of the five *degrees* (~ 12 min $\times 5$):
- #4.1 The **EMG calibration** was performed.
- #4.2 A first **learning test trial** was performed.
- #4.3 The participants performed the **regular exercise** and the **exercise with the secondary task**. For D1 and D3, the **exercise with the auditory privation** was also carried out. The order of all three exercises was systematically permuted and after each one a **perception questionnaire** was filled out.

The participants performed altogether 5 (regular, D1–D5) + 5 (secondary task, D1–D5) + 2 (auditory privation, D1 and D3) = 12 exercises. As for the perception questionnaire, it comprised items related to the *perception of realism* (seven-point Likert scale) and *workload* (1–20 scale, reported in five quantiles in the results for the sake of clarity: Q1 – Q5). Regarding the *perception of realism*, the participants had to evaluate how realistic the **overall** experience was, how realistic the **contacts** felt, and how realistic the **manipulation** or movement of the objects in the scene was. As for the *workload*, the **physical** and the **mental** effort had to be rated; as explained to the participants during the instructions, the first is related to the corporal ease when exerting movements, whereas the second comprises understanding and planning strategies for optimally fulfilling the exercises. Longer standardized questionnaires were avoided to reduce session time.

In addition to the subjective ratings, the following objective values were recorded at 50 Hz: trajectory along time, real and virtual contact forces and torques, and muscular effort signals; for exercises with the *secondary task*, reaction times were also measured with 1 kHz accuracy, in addition to the missed signals.

5 RESULTS AND DISCUSSION

Results are presented and discussed in two main sections: Sec. 5.1 deals with the regular exercises, whereas Sec. 5.2 covers exercises with the *secondary task* and the *auditory privation*. Especial focus is put on the set of regular exercises. The most important insights are enumerated and discussed in Sec. 5.3; all items in it are linked throughout the text (e.g., \rightarrow L01).

The statistical analysis was performed in R using standard packages and the `ez` library [32]. For the objective continuous variables, parametric tests were used. All objective variables were normally distributed in most of the cases; it is trusted in the robustness of the parametric methods for the exceptions. Additionally, sphericity corrections were applied when necessary (values given in tables). On the other hand, non-parametric tests (less powerful, but more appropriate for discrete values) were used for the subjective data. In all cases, the main significance level was set to $\alpha = 0.05$, but this value was adjusted with the Bonferroni factor b for pairwise comparisons (i.e., $\alpha' = \alpha/b = 0.05/10 = 0.005$). For the sake of brevity and clarity, it is avoided introducing too many statistical values in the text and the reader is encouraged to look up in the provided tables.

TABLE 1

Descriptive data of the objective dependent variables for the whole exercise and the tasks. The five *Degrees* (D) are decomposed in the three varied factors *Visual Feedback* (VF), *Haptic Device* (HD), and *Haptic Rendering* (HR). The used *real* (R) or *synthetic* (S) treatment of the factors is also specified (see Fig. 5). Average and standard deviation values are provided for each *Degree* (D), as well as bar diagrams of the total values. The statistical analysis can be found in Table 2.

D	VF	HD	HR	Total Exercise	Tasks		
					Frontal	Side	Railing
(a) Time to Complete [s]							
D1	R	R	R	15.76 (3.50)	4.26 (1.12)	5.60 (1.30)	5.89 (1.39)
D2	S	R	R	21.60 (4.75)	5.81 (1.56)	7.81 (1.78)	7.98 (1.67)
D3	R	S	R	31.74 (6.43)	7.26 (1.53)	12.59 (2.53)	11.89 (2.85)
D4	S	S	R	38.54 (8.51)	9.11 (1.84)	15.02 (3.04)	14.41 (4.34)
D5	S	S	S	55.27 (16.02)	12.20 (4.46)	22.12 (9.69)	20.95 (8.55)
<i>Grand M & SD</i>				32.58 (16.46)	7.73 (3.65)	12.63 (7.49)	12.23 (6.95)
(b) Average Force [N]							
D1	R	R	R	1.06 (0.53)	0.87 (0.34)	0.91 (0.27)	1.26 (0.94)
D2	S	R	R	1.08 (0.59)	0.95 (0.53)	0.98 (0.35)	1.25 (0.94)
D3	R	S	R	1.45 (0.74)	1.13 (0.56)	1.18 (0.52)	1.79 (1.26)
D4	S	S	R	1.76 (0.62)	1.72 (0.78)	1.65 (0.72)	1.83 (0.98)
D5	S	S	S	6.70 (2.95)	6.01 (3.72)	6.87 (2.82)	6.07 (3.18)
<i>Grand M & SD</i>				2.41 (2.59)	2.14 (2.61)	2.32 (2.65)	2.44 (2.48)
(c) Average Effort [0: relaxation, 100: steady maximum during calibration activity]							
D1	R	R	R	56.88 (14.13)	53.24 (14.91)	61.59 (16.26)	55.27 (14.68)
D2	S	R	R	54.03 (16.83)	51.00 (18.32)	55.46 (16.82)	54.90 (17.93)
D3	R	S	R	90.13 (37.23)	86.91 (40.46)	95.39 (42.03)	86.62 (31.38)
D4	S	S	R	91.48 (47.02)	86.63 (42.89)	95.47 (49.80)	90.77 (47.87)
D5	S	S	S	117.75 (39.63)	117.48 (41.51)	118.14 (35.48)	115.69 (47.80)
<i>Grand M & SD</i>				82.05 (40.82)	79.05 (41.57)	85.21 (41.50)	80.65 (41.44)

5.1 Regular Exercises

Table 1 gathers the most relevant descriptive data of the objective dependent variables, whereas Table 2 summarizes the results of their statistical analysis. In Table 1, the task segment values of the objective variables are shown, but only the total exercise values are considered for the statistical analysis. As far as the subjective variables are concerned, Table 3 outlines the descriptive data and Table 4 their statistical analysis.

A repeated measures one-way ANOVA was performed on the **Degree of Virtualization** (D) for total exercise values of both objective and subjective dependent variables and it turned out to be significant in all cases (all $ps < 0.001$, ***). Therefore, the following subsections will deal with the post hoc pairwise comparisons between the degrees to discuss the effect of the HMD¹¹ (VF: D1–D2, D3–D4), the HUG (HD: D1–D3), the haptic rendering method (HR: D4–D5), and the whole system (VR: D1–D5).

5.1.1 Objective Variables: Performance (Table 1 & Table 2)

After a general visual inspection, the bar diagrams (Table 1) show that the *Time to Complete* progressively increases with the **Degree** (increment ratios varying between $\sim 20 - 50\%$). In the case of the *Average Force*, the *synthetic* values (D5) are considerably much higher (up to $\sim 6\times$) than the *real* ones (D1–D4). And, finally, the *Average Effort* substantially

11. When the head-mounted display (HMD) is mentioned, the whole *synthetic* visual feedback is meant, including the graphical rendering.

TABLE 2

Statistical analysis of the objective dependent variables (performance) to determine the effect of the *Degree* (One-Way ANOVA with Fisher tests) and the factors of *Visual Feedback* (VF), *Haptic Device* (HD), and *Haptic Rendering* (HR). Pairwise comparisons were done with T-Tests after a Bonferroni adjustment of $b = 10$. Provided values: Sphericity correction (Mauchly’s W and Greenhouse-Geisser ϵ), sample size (N), degrees of freedom (df), S statistic (S is Fisher $F(df, N - 1)$ for *Degree* and T-Test $T(df)$ for pairwise comparisons), p -value, Cohen’s d , and the relation (with coded effect size) between the *degrees*. The source descriptive data can be found in Table 1.

	W	$p(>W)$	ϵ	b	N	df	S	$p(>S)$	sig.	d	Relation
(a) Time to Complete (Total Exercise)											
Degree	4.86e-02	1.88e-10	0.41	(1)	24	4	112.57	1.40e-15	***	-	-
D1-D2 (VF)	-	-	-	10	24	23	-8.11	3.38e-08	***	1.40	D2 \gg D1
D1-D3 (HD)	-	-	-	10	24	23	-15.71	8.68e-14	***	3.08	D3 \gg D1
D3-D4 (VF)	-	-	-	10	24	23	-5.67	8.97e-06	***	0.90	D4 \gg D3
D4-D5 (HR)	-	-	-	10	24	23	-5.94	4.74e-06	***	1.30	D5 \gg D4
D1-D5 (VR)	-	-	-	10	24	23	-12.98	4.58e-12	***	3.41	D5 \gg D1
(b) Average Force on Contact (Total Exercise)											
Degree	8.45e-04	8.34e-28	0.28	(1)	24	4	79.03	1.32e-09	***	-	-
D1-D2 (VF)	-	-	-	10	24	23	-0.30	0.77		0.05	D2 \approx D1
D1-D3 (HD)	-	-	-	10	24	23	-4.18	3.54e-04	**	0.62	D3 $>$ D1
D3-D4 (VF)	-	-	-	10	24	23	-3.26	3.41e-03	*	0.46	D4 \gtrsim D3
D4-D5 (HR)	-	-	-	10	24	23	-8.43	1.72e-08	***	2.31	D5 \gg D4
D1-D5 (VR)	-	-	-	10	24	23	-9.53	1.87e-09	***	2.66	D5 \gg D1
(c) Average Effort (Total Exercise)											
Degree	0.145	4.52e-06	0.70	(1)	24	4	24.90	2.10e-10	***	-	-
D1-D2 (VF)	-	-	-	10	24	23	-1.33	0.20		0.18	D2 \approx D1
D1-D3 (HD)	-	-	-	10	24	23	-4.89	6.19e-05	***	1.18	D3 \gg D1
D3-D4 (VF)	-	-	-	10	24	23	-0.16	0.87		0.03	D4 \approx D3
D4-D5 (HR)	-	-	-	10	24	23	-2.71	1.25e-02	.	0.60	D5 $>$ D4
D1-D5 (VR)	-	-	-	10	24	23	-9.37	2.56e-09	***	2.05	D5 \gg D1

Significance codes (p): 0 *** 0.001/b ** 0.01/b * 0.05/b . 0.1/b 1

Effect size codes (d): 0 \approx 0.2 \gtrsim 0.5 $>$ 0.8 \gg 1.5 \ggg ∞

augmentations ($\sim 60\%$) when the haptic device comes into play (D3 onwards).

All pairwise comparisons related to the *Time* values are significant (all $ps < 1e-04$, ***). The HMD increased the means $+37\%$ when used alone (D1-D2) and $+21\%$ when used in combination with the HUG (D3-D4). The HUG, in addition, produced $2\times$ larger values (D1-D3), and the haptic rendering method led to $+43\%$ bigger recordings. Overall, the participants needed $3.5\times$ the *Time* of the purely real **Degree** (D1) to complete the exercises in the purely virtual one (D5) (\rightarrow L01). Interestingly, the effects of the *synthetic* treatments do not seem to add up linearly. Additionally, the haptic feedback, and in particular, the haptic device HUG seems to be the most important bottleneck when it comes to the *Time* performance.

Post hoc pairwise comparisons of *Force* values revealed distinct difference levels between **Degrees**. The HMD did not significantly affect the *Forces* exerted by the participants when used alone (D1-D2, $p > 0.005$, ns), but it contributed to $+22\%$ larger means when used in combination with the HUG (D3-D4, $p < 0.005$, *). The HUG alone significantly contributed to $+38\%$ bigger values (D1-D3, $p < 0.001$, **). Moreover, *Synthetic* contact rendering led to $3.79\times$ higher *Force* values (D4-D5, $p < 1e-04$, ***). When it comes to the objective indicators, it is in the *Forces* where the purely virtual **Degree** (D5) most differentiates from the purely real one (D1), with $6.35\times$ bigger means for the *synthetic* condition (D1-D5, $p < 1e-04$, ***) (\rightarrow L02). Clearly, these

virtual *Forces* are too high compared to the real ones. Decreasing the value of the stiffness could seem to be a way of solving that, however, as mentioned in Sec. 4.1, the used stiffness (lower than the maximum possible) was selected to produce the contacts with the highest fidelity possible. Due to the observational evidence and the insights from previous studies [18], the reason for such high *Forces* could be that the haptic rendering algorithm causes interactions that are more jamming than real ones; as a result, the user applies higher *Forces* and requires also longer to accomplish the exercises. This point should be further investigated.

No significant effect of the HMD or the haptic rendering was found on the *Effort* values (all $ps > 0.005$, ns). However, the HUG led to significantly $+58\%$ bigger values (D1-D3, $p < 1e-04$, ***). In the same line, muscular *Effort* signals were on average $2\times$ larger in the purely virtual **Degree** compared to the purely real (D1-D5, $p < 1e-04$, ***) (\rightarrow L03). To the best of our knowledge, no other comparison studies of this sort have used EMG measurements as an indicator. The found significant differences proof the validity of it.

5.1.2 Subjective Variables: Perception (Table 3 & Table 4)

A first look at the histograms (Table 3) conveys the fact that the *Overall*, *Contact*, and the *Manipulation Realism* ratings progressively shift from maximum values (7) in D1 to values gravitating around 4 – 5 in D5. In contrast, the *Physical* and *Mental Workload* seem to increase with the **Degree** from quantiles Q1–Q2 (ratings 1 – 8 of a maximum of 20) to quantiles Q2–Q3 (ratings 5 – 12).

The three ratings of *Realism Perception* decreased roughly 1 point when the HMD was used (alone) in interaction with the real models (D1-D2, all $ps < 0.005$, or smaller levels). However, no significant differences could be found due to the HMD when it was used in combination with the HUG (D3-D4, all $ps > 0.005$). The use of the HUG yielded ratings which were 1 – 2 points lower for the *Overall* and the *Manipulation Realism* (D1-D3, both $ps < 1e-04$, ***); the *Contact Realism* decreased roughly 1 point ($p < 0.001$, **). Regarding the haptic rendering method, a significant reduction of 1 point was found only in the case of the *Contact Realism* (D4-D5, $p < 0.001$, **) (\rightarrow L04). Comprehensibly, it seems that the HMD affects primarily the *Overall Perception of Realism*, the haptic rendering method the *Contact Realism*, and the HUG the *Manipulation Realism*; moreover, the HUG and the haptic rendering method are the *synthetic* treatments with the highest impact.

Both *Physical* and *Mental Workload* average ratings of the purely virtual condition (D5) were respectively $2\times$ ($p < 1e-04$, ***) and $3\times$ ($p < 0.001$, **) higher than the purely real ones (D1). Nevertheless, the *synthetic* treatments affected differently each dimension. In the case of the HMD, a significant increase of $+1$ average point could be determined for the *Physical Workload* when used alone, while the same significant increment occurred when it was used in combination with the HUG for the *Mental Workload* (both $ps < 0.005$, *). The HUG led to $2.21\times$ higher *Physical Workload* ratings ($p < 1e-04$, ***), but no significant effect could be found on the *Mental Workload*. Regarding the haptic rendering algorithm, slight significances were found for both the *Physical* ($p < 0.005$, *) and the *Mental Workload* ($p < 0.05$, .); the *synthetic* treatment resulted in ratings $+2$

TABLE 4

Statistical analysis of the subjective dependent variables to determine the effect of the *Degree* (One-Way Friedman ANOVA) and the factors of *Visual Feedback* (VF), *Haptic Device* (HD), and *Haptic Rendering* (HR). Pairwise comparisons were done with T-Tests after a Bonferroni adjustment of $b = 10$. Provided values: Sample size (N), degrees of freedom (df), S statistic (S is Friedman χ^2 (df) for *Degree* and Wilcoxon V for pairwise comparisons), p -value, Cliff’s δ , and the relation (with coded effect size) between the *degrees*. The source descriptive data can be found in Table 3.

	N	df	b	S	$p(>S)$	sig.	δ	Relation
(a) Overall Realism (Total Exercise)								
Degree	24	4	(1)	68.58	4.53e-14	***	-	-
D1–D2 (VF)	24	23	10	190	6.33e-05	***	0.79	D1 \ggg D2
D1–D3 (HD)	24	23	10	231	4.55e-05	***	0.87	D1 \ggg D3
D3–D4 (VF)	24	23	10	98.5	0.107		0.26	D3 \gtrsim D4
D4–D5 (HR)	24	23	10	86.5	2.81e-02		0.32	D4 $>$ D5
D1–D5 (VR)	24	23	10	300	1.19e-05	***	0.97	D1 \ggg D5
(b) Contact Realism (Total Exercise)								
Degree	24	4	-	66.19	1.45e-13	***	-	-
D1–D2 (VF)	24	23	10	55	1.90e-03	*	0.42	D1 $>$ D2
D1–D3 (HD)	24	23	10	120	4.56e-04	**	0.63	D1 \gg D3
D3–D4 (VF)	24	23	10	56	0.842		0.01	D3 \approx D4
D4–D5 (HR)	24	23	10	226	1.07e-04	**	0.73	D4 \ggg D5
D1–D5 (VR)	24	23	10	300	1.57e-05	***	1	D1 \ggg D5
(c) Manipulation Realism (Total Exercise)								
Degree	24	4	(1)	66.53	1.22e-13	***	-	-
D1–D2 (VF)	24	23	10	91	9.09e-04	**	0.54	D1 \gg D2
D1–D3 (HD)	24	23	10	253	3.61e-05	***	0.92	D1 \ggg D3
D3–D4 (VF)	24	23	10	59	0.40		0.08	D3 \approx D4
D4–D5 (HR)	24	23	10	125.5	1.89e-02		0.35	D4 $>$ D5
D1–D5 (VR)	24	23	10	300	1.63e-05	***	1	D1 \ggg D5
(d) Physical Workload (Total Exercise)								
Degree	24	4	(1)	72.66	6.21e-15	***	-	-
D1–D2 (VF)	24	23	10	4.5	1.42e-03	*	0.24	D2 \gtrsim D1
D1–D3 (HD)	24	23	10	0	4.09e-05	***	0.70	D3 \ggg D1
D3–D4 (VF)	24	23	10	45	0.137		0.07	D4 \approx D3
D4–D5 (HR)	24	23	10	14	1.70e-03	*	0.26	D5 \gtrsim D4
D1–D5 (VR)	24	23	10	0	1.89e-05	***	0.85	D5 \ggg D1
(e) Mental Workload (Total Exercise)								
Degree	24	4	(1)	46.63	1.82e-09	***	-	-
D1–D2 (VF)	24	23	10	29	1.42e-02		0.32	D2 \gtrsim D1
D1–D3 (HD)	24	23	10	65	2.67e-02		0.33	D3 \gtrsim D1
D3–D4 (VF)	24	23	10	36	3.19e-03	*	0.24	D4 \gtrsim D3
D4–D5 (HR)	24	23	10	27	6.18e-03	.	0.32	D5 $>$ D4
D1–D5 (VR)	24	23	10	10.5	1.10e-04	**	0.75	D5 \ggg D1

Significance codes (p):

0 *** 0.001/ b ** 0.01/ b * 0.05/ b . 0.1/ b 1

Effect size codes (δ):

0 \approx 0.15 \gtrsim 0.3 $>$ 0.5 \gg 0.7 \ggg 1

- L05** The *Perception of Physical and Mental Workload* shifted roughly from a 20% level in the purely real condition (D1) to a 50% level in the purely virtual one (D5); the HMD and the haptic rendering contributed approximately to increments of 5–10% points each, whereas the HUG added 20% points, especially in the case of the *Physical Workload* (Sec. 5.1.2).
- L06** During the exercises with the secondary task, no significant effect of the **Degree** on the *Reaction Times* could be found; however, the number missed horn signals significantly correlated with the *Overall Realism* ($\rho = -0.25$) and the *Mental Workload* ($\rho = 0.25$) (Sec. 5.2.1).
- L07** No changes in the performance could be determined when contact sounds were silenced; however, almost half of the participants rated the *Overall Realism* 1 point smaller (out of 7) in the purely real

condition wearing headphones (D1) (Sec. 5.2.2).

As expected, subjects showed better performance during purely real exercises (D1) than in purely virtual ones (D5); along these lines, objective indicators show that, compared to *Degree* D1, *Degree* D5 led to $3.51\times$ larger *Time* values (L01), $6.35\times$ larger *Force* values (L02), and $2\times$ larger *Effort* values (L03). However, the contribution of the work lies rather on the granular quantification of disparities produced by each synthetic feedback (subsystem) on the analyzed objective and subjective indicators.

In general, haptic feedback seems to be the bottleneck. Additionally, as far as synthetic haptic feedback is concerned, two noteworthy facts that account for most of the variance between real and virtual manipulations can be observed: (i) the HUG had the highest impact on the *Time* (L01), *Effort* (L03), and *Physical Workload* (L04) indicators, and (ii) the *synthetic* forces were considerably higher than the *real* ones (L02).

It is worth to mention that, overall, *Time* ratios between real and virtual manipulations (L01) were close to the ones reported in some works cited in Sec. 2: [12], [14], and [15]. In contrast to this present study, those related works employed desktop-size haptic interfaces. Our previous evaluation [18] concluded that around 20% shorter *Time* values could be achieved using the Sigma.7 desktop device, although a scaling of $5.5\times$ was necessary to cover the whole virtual workspace. Along these lines, smaller *Times* should be expected if smaller and lighter devices are used. Additionally, the notable increase of the *Average Muscular Effort* (L03) and *Physical Workload* (L05) due to the HUG seems to be related to that increase of *Time* values. As noted in [18], the HUG seems to increase contact realism perception, but in detriment of ergonomics ratings. The device is a remarkable research platform for testing large unscaled movements and a broad stiffness spectrum; nonetheless, the results seem to point to be the optimum device one with a similar large workspace, but which sacrifices stiffness capabilities for lower inertia values.

The unexpectedly large virtual forces (L02) reveal an improvement direction for the haptic rendering algorithm. However, it is worth to mention that the commanded virtual forces (D5) analyzed here differ from the actual exerted forces, since the influence of the robot dynamics and the hand-device coupling are not contemplated; yet, for practical reasons, it is assumed both are similar. In contrast to our results, Unger et al. [12] found no significant differences between real and virtual forces in peg-in-hole scenarios using a desktop-size magnetic levitation haptic interface and the Coriolis engine [33] for force rendering. After observing recorded exercises of D5 and conducting short interviews with the participants, two major interpretations that help explain that discrepancy between force magnitudes were gathered: (i) as opposed to other penalty-based haptic rendering algorithms, the used constraint-based approach (Sec. 3.2) tends to sharpen all edges, sometimes even reducing the natural slipperiness of small corner contacts, and (ii) the perceived ratio between torques and forces transmitted (i.e., not commanded) through the HUG might be smaller than in the real world. Therefore, users could have been provided with realistic but sometimes misleading collision

cues, especially during insertion tasks in which accurate forces are necessary to comprehend the configuration error and perform fast corrections. Due to that resulting lower fidelity, participants might have literally *worked hard and not smart*, as opposed to how the saying goes. Our previous study [18] also supports these points; nevertheless, further investigation is required to substantiate the interpretation. Along these lines, and following the debate in the VR community discussing to what extent immersion components should mimic reality [34], further research questions worth exploring arise: Which is the minimum fidelity necessary to obtain the maximum skill transfer? How do fidelity improvements affect performance?

Regarding the *Realism* ratings, each *synthetic* treatment seems to contribute to steady, slow, and similar decreases (L04). In this sense, the *Perception of Realism* is probably a more robust and less sensible variable than the objective indicators.

As far as the series of exercises beyond the regular one are concerned, more modest effects were found; yet, this reflects the effectiveness and influence of the methods and modalities tested in them, which is also a finding. The impact and explanatory power of the secondary task (L06) are lower than the ones suggested in the literature [28], [29]. Thus, reactions to parallel tasks could not be as explanatory when studying assembly manipulations similar to the ones tested in this work. Similarly, although the privation of auditory cues slightly decreased the *Overall Perception Realism*, no effects were found in the performance indicators (L07); however, this is in line with the findings from Gupta et al. [5]. Sounds related to harder materials have been shown to bias the perception of stiffness towards higher ratings [30], [31], yet real audio cues seem not to be as dominant for the analyzed variables in the tested scenarios. Therefore, synthetic sound could probably have a secondary priority compared to haptic feedback in virtual assembly manipulations.

6 CONCLUSIONS AND PERSPECTIVES

Although virtual manipulation with haptic feedback has been proven to be a useful tool in many fields, their fidelity is still noticeably moderate when compared to purely real interactions; this might hinder the practical use of virtual setups in training and planning applications. Literature suggests that satisfactory perceptions and performances are the result of the interplay between different feedback modalities and their constituent blocks. Along these lines, we performed a user evaluation in which $N = 24$ participants performed well-motivated assembly exercises in five different conditions, ranging gradually from purely real to purely virtual, and systematically introducing different feedback devices (i.e., haptic device or head-mounted display) or virtual feedback signals (i.e., haptic feedback). All participants tried all conditions in a different order. This work is the continuation of a previous one [18] in which the effects of different haptic devices and haptic rendering algorithms on user performance were analyzed.

On the other hand, this current study sheds light on the understanding of the difference between the real and virtual manipulations by analyzing the effects of each system component. Moreover, altogether *visual*, *haptic*, and

audio feedback modalities were considered, leading to a multi-modal perspective to interpret the formation of the virtual contact percept. While previous works have mainly analyzed task completion time performances, we provide with more than three objective (*Time*, *Force*, and *Effort*) and five subjective (related to *Realism* and *Workload*) dimensions for each tested condition. To the best of our knowledge, this is the first multimodal and system-wise study covering so many variables. Additionally, the results are easily replicable and transferable thanks to the used scenario, consisting of a large workspace allowing unscaled movements and well-motivated but abstract exercises.

Previous Sec. 5.3 synthesizes and discusses the most important insights, and their relationship to the literature. Independently of the systems involved, virtual manipulations inevitably lead worse performance and subjective results compared real ones; nonetheless quantifying them properly is an imperative step. In general, the haptic modality seems to be the bottleneck for performance: HUG, the used haptic device, affects the most *Completion Time* and *Muscular Effort* indicators; in the same line, synthetically rendered forces differ from real ones significantly, although the effect of that needs to be studied more deeply. On the other hand, the decrease of *Realism Perception* caused by any synthetic device or rendering is similar and sound seems not to be as relevant.

The preceding study [18] suggested that a haptic device with a large upper-body workspace but moderate stiffness capacity in favor of a lighter structure could significantly improve both performance and perception indicators. That conclusion opposes to the mainstream desktop-size systems and requires an improvement from the HUG itself. In addition to showing that the effects of the HUG on performance and subjective variables are the largest, the current study thoroughly quantifies those effects and contextualizes them in relation to other multimodal factors. That helps to a more holistic understanding.

Short term future work will deal with investigating and enhancing the two main bottlenecks detected in this work: (i) the effect of the HUG on the user and (ii) the discrepancy between physical and computed contact forces. Additionally, we will try to formalize result mappings between setups of similar and different features performing further comparison studies. Our ultimate goal is to study and improve the skill learning and transfer through virtual environments, focusing on practical real-world scenarios. In particular, we are interested in interactive virtual assembly [24] and maintenance [35] applications.

REFERENCES

- [1] T. Coles, D. Meglan, and N. John, “The role of haptics in medical training simulators: A survey of the state of the art,” *IEEE Trans. on Haptics*, vol. 4, no. 1, pp. 51–66, 2011.
- [2] P. Xia, “Haptics for product design and manufacturing simulation,” *IEEE Trans. on Haptics*, vol. 9, no. 3, pp. 358–375, 2016.
- [3] M. Sagardia, B. Weber, T. Hulin, C. Preusche, and G. Hirzinger, “Evaluation of visual and force feedback in virtual assembly verifications,” in *Proc. IEEE Virtual Reality (VR)*. IEEE, 2012, pp. 23–26.
- [4] B. Weber, M. Sagardia, T. Hulin, and C. Preusche, “Visual, vibrotactile, and force feedback of collisions in virtual environments: effects on performance, mental workload and spatial orientation,” in *Virtual Augmented and Mixed Reality. Designing and Developing*

- Augmented and Virtual Environments*, ser. Lecture Notes in Computer Science, R. Shumaker, Ed. Springer Berlin Heidelberg, 2013, vol. 8021, pp. 241–250.
- [5] R. Gupta, T. Sheridan, and D. Whitney, "Experiments using multimodal virtual environments in design for assembly analysis," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 3, pp. 318–338, 1997.
 - [6] S. Garbaya and U. Zaldivar-Colado, "The affect of contact force sensations on user performance in virtual assembly tasks," *Virtual Reality*, vol. 11, no. 4, pp. 287–299, 2007.
 - [7] L. B. Rosenberg and B. D. Adelstein, "Perceptual decomposition of virtual haptic surfaces," in *Proc. IEEE Virtual Reality (VR)*. IEEE, 1993, pp. 46–53.
 - [8] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, 2002.
 - [9] D. Hecht and M. Reiner, "Sensory dominance in combinations of audio, visual and haptic stimuli," *Experimental brain research*, vol. 193, no. 2, pp. 307–314, 2009.
 - [10] T. Hulin, K. Hertkorn, P. Kremer, S. Schätzle, J. Artigas, M. Sagardia, F. Zacharias, and C. Preusche, "The DLR bimanual haptic device with optimized workspace (video)," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*. IEEE, 2011, pp. 3441–3442.
 - [11] M. Sagardia and T. Hulin, "A fast and robust six-dof god object heuristic for haptic rendering of complex models with friction," in *Proc. ACM Symp. on Virtual Reality and Software Technology (VRST)*. ACM, 2016, pp. 163–172.
 - [12] B. J. Unger, A. Nicolaidis, A. Thompson, R. L. Klatzky, R. L. Hollis, P. J. Berkelman, and S. Lederman, "Virtual peg-in-hole performance using a 6-dof magnetic levitation haptic device: Comparison with real forces and with visual guidance alone," in *Proc. IEEE Haptics Symposium (HAPTICS)*. IEEE, 2002, pp. 263–270.
 - [13] A. M. Bashir, R. Bicker, and P. M. Taylor, "An investigation into different visual/tactile feedback modes for a virtual object manipulation task," in *Proc. ACM SIGGRAPH Int. Conf. on Virtual Reality Continuum and its Applications in Industry*. ACM, 2004, pp. 359–362.
 - [14] T. Lim, J. M. Ritchie, R. G. Dewar, J. R. Corney, P. Wilkinson, M. Calis, M. Desmulliez, and J.-J. Fang, "Factors affecting user performance in haptic assembly," *Virtual Reality*, vol. 11, no. 4, pp. 241–252, 2007.
 - [15] G. Gonzalez-Badillo, H. Medellin-Castillo, T. Lim, J. Ritchie, and S. Garbaya, "The development of a physics and constraint-based haptic virtual assembly system," *Assembly Automation*, vol. 34, no. 1, pp. 41–55, 2014.
 - [16] M. Harders, A. Barlit, K. Akahane, M. Sato, and G. Szekely, "Comparing 6dof haptic interfaces for application in 3d assembly tasks," in *Proc. Eurohaptics Conf.*, vol. 6, 2006, pp. 523–526.
 - [17] E. Samur, F. Wang, U. Spaelter, and H. Bleuler, "Generic and systematic evaluation of haptic interfaces based on testbeds," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2007, pp. 2113–2119.
 - [18] M. Sagardia and T. Hulin, "Evaluation of a penalty and a constraint-based haptic rendering algorithm with different haptic interfaces and stiffness values," in *Proc. IEEE Virtual Reality (VR)*. IEEE, 2017, pp. 64–73.
 - [19] F. Zacharias, I. S. Howard, T. Hulin, and G. Hirzinger, "Workspace comparisons of setup configurations for human-robot interaction," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2010, pp. 3117–3122.
 - [20] P. Kremer, T. Wimböck, J. Artigas, S. Schätzle, K. Jöhl, F. Schmidt, C. Preusche, and G. Hirzinger, "Multimodal telepresent control of DLR's Rollin' JUSTIN (video)," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*. IEEE, 2009, pp. 1601–1602.
 - [21] M. Sagardia, T. Stouraitis, and J. L. e Silva, "A New Fast and Robust Collision Detection and Force Computation Algorithm Applied to the Physics Engine Bullet: Method, Integration, and Evaluation," in *Prof. of the Conf. and Exhibition of the European Association of Virtual and Augmented Reality (EuroVR)*. Eurographics Association, 2014, pp. 65–76.
 - [22] W. A. McNeely, K. D. Puterbaugh, and J. J. Troy, "Six degree-of-freedom haptic rendering using voxel sampling," in *Proc. ACM SIGGRAPH*. ACM, 1999, pp. 401–408.
 - [23] H. Xu and J. Barbic, "Adaptive 6-dof haptic contact stiffness using the gauss map," *IEEE Trans. on Haptics*, vol. 9, no. 3, pp. 323–332, 2016.
 - [24] M. Sagardia, T. Hulin, K. Hertkorn, P. Kremer, and S. Schätzle, "A platform for bimanual virtual assembly training with haptic feedback in large multi-object environments," in *Proc. ACM Symp. on Virtual Reality and Software Technology (VRST)*. ACM, 2016, pp. 153–162.
 - [25] J. Wildenbeest, D. Abbink, C. Heemskerk, F. van der Helm, and H. Boessenkool, "The impact of haptic feedback quality on the performance of teleoperated assembly tasks," *IEEE Trans. on Haptics*, vol. 6, no. 2, pp. 242–252, 2013.
 - [26] A. Bloomfield, Y. Deng, J. Wampler, P. Rondot, D. Harth, M. McManus, and N. Badler, "A taxonomy and comparison of haptic actions for disassembly tasks," in *Proc. IEEE Virtual Reality (VR)*. IEEE, 2003, pp. 225–231.
 - [27] B. Deml, "Human factors issues on the design of telepresence systems," *Presence: Teleoperators and Virtual Environments*, vol. 16, no. 5, pp. 471–487, 2007.
 - [28] K. Hertkorn, "Shared grasping: A combination of telepresence and grasp planning," Ph.D. dissertation, Karlsruher Institut für Technologie (KIT) and German Aerospace Center (DLR), 2015.
 - [29] C. C. Bracken, G. Pettey, and M. Wu, "Revisiting the use of secondary task reaction time measures in telepresence research: exploring the role of immersion and attention," *AI & society*, vol. 29, no. 4, pp. 533–538, 2014.
 - [30] D. E. DiFranco, G. L. Beauregard, and M. A. Srinivasan, "The effect of auditory cues on the haptic perception of stiffness in virtual environments," in *Proc. of the ASME Dynamic Systems and Control Division*, vol. 61. American Society of Mechanical Engineers (ASME), 1997, pp. 17–22.
 - [31] F. Avanzini and P. Crosato, "Haptic-auditory rendering and perception of contact stiffness," in *Int. Workshop on Haptic and Audio Interaction Design*. Springer, 2006, pp. 24–35.
 - [32] M. A. Lawrence, *ez: Easy analysis and visualization of factorial experiments*, 2011. [Online]. Available: <http://CRAN.R-project.org/package=ez>
 - [33] D. Baraff, "Issues in computing contact forces for non-penetrating rigid bodies," *Algorithmica*, vol. 10, no. 2-4, pp. 292–352, 1993.
 - [34] D. A. Bowman and R. P. McMahan, "Virtual reality: how much immersion is enough?" *Computer*, vol. 40, no. 7, pp. 36–43, 2007.
 - [35] M. Sagardia, K. Hertkorn, T. Hulin, S. Schätzle, R. Wolff, J. Hummel, J. Dodiya, and A. Gerndt, "VR-OOS: The DLR's virtual reality simulator for telerobotic on-orbit servicing with haptic feedback," in *Proc. IEEE Aerospace Conf.*, 2015, pp. 1–17.



model-mediated telemanipulation.

Mikel Sagardia studied Mechanical Engineering at University of Navarra (Tecnun, Spain) and at the Technical University of Munich (TUM, Germany), graduating with a M.Sc. degree in 2008. He works since then at the Institute of Robotics and Mechatronics of the German Aerospace Center (DLR). As of late 2013, he is also an external PhD candidate at the TUM. His research interests include collision detection and physical simulation, haptic rendering, human and machine haptics perception, virtual reality and



Thomas Hulin received his Dipl.-Ing. degree from the Technical University of Munich in 2003. In that year, he also joined the Institute of Robotics and Mechatronics, German Aerospace Center (DLR), Wessling, Germany. In early 2017, he received his Dr.-Ing. degree from the Leibniz University of Hannover. His research interests include haptic devices, control and algorithms, physical human-robot interaction, tele- and space robotics, robot visualization, virtual reality, and skill transfer.