Design and Evaluation of a Handheld-based 3D User Interface for Collaborative Object Manipulation

Jerônimo Gustavo Grandi[1], Henrique Galvan Debarba[23], Luciana Nedel[1] and Anderson Maciel[1]
[1]Federal University of Rio Grande do Sul, Institute of Informatics
[2]École Polytechnique Fédérale de Lausanne, Immersive Interaction Group
[3]Artanim Foundation, Geneva, Switzerland
{jggrandi,nedel,amaciel}@inf.ufrgs.br, henrique.debarba@artanim.ch

ABSTRACT
Object manipulation in 3D virtual environments demands a combined coordination of rotations, translations and scales, as well as the camera control to change the user's viewpoint. Then, for many manipulation tasks, it would be advantageous to share the interaction complexity among team members. In this paper we propose a novel 3D manipulation interface based on a collaborative action coordination approach. Our technique explores a smartphone – the touchscreen and inertial sensors – as input interface, enabling several users to collaboratively manipulate the same virtual object with their own devices. We first assessed our interface design on a docking and an obstacle crossing tasks with teams of two users. Then, we conducted a study with 60 users to understand the influence of group size in collaborative 3D manipulation. We evaluated teams in combinations of one, two, three and four participants. Experimental results show that teamwork increases accuracy when compared with a single user. The accuracy increase is correlated with the number of individuals in the team and their work division strategy.

ACM Classification Keywords
H.5.2. User Interfaces: Input devices and strategies; H.5.3. Group and Organization Interfaces: Computer-supported cooperative work

Author Keywords
3D User Interfaces; Collaborative Manipulation; User Studies.

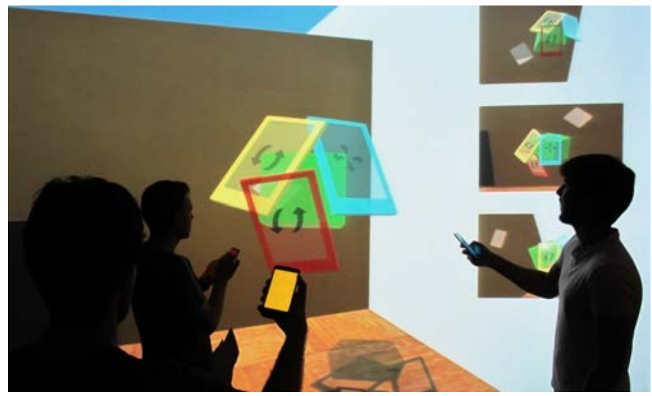


Figure 1: Three users simultaneously driving the object through the virtual environment. The colored rectangles indicate the position and orientation of each user in the VE. The three windows at the right-side show the three personal views.

INTRODUCTION
The accomplishment of spatial tasks in 3D virtual environments (VE) involves a complex coordination of virtual manipulations, such as translation, rotation and scale [27]. Each of these transformations refers to independent degrees of freedom (DOFs). While these manipulations are fundamental to any 3D virtual environment, performing them freely demands a highly cognitive effort to coordinate the possible transformations [25]. 3D user interfaces (3DUI) try to reduce the mental overhead by mapping natural gestures and movements into the corresponding action in the VE. However, create effective interfaces for 3D virtual systems is challenging [24]. Many 3DUI techniques have been proposed to allow natural, precise and fast interaction with virtual environments [5]. Techniques are typically designed for specific purpose applications. Some of them have become relatively common due to the easy access to the market, such as gaming console interfaces. Nevertheless, as of today, research in the 3D object manipulation field mainly focus on single-user interaction.

One alternative towards 3D manipulations efficiency is to share the work and solve the task in a parallel and collaborative manner. Collaborators may wish to split the different aspects of the manipulation among them. Thus, perhaps, accomplishing the assignment faster, more accurately and with less fatigue. Teamwork, in turn, involves considerable negotiations [6] and, as team members vary, team strategies and task accomplishment processes change as well. Modeling such human interactions raises novel collaborative concepts compared to those typically grounded in a single-user scenario [19]. An effective design of a collaborative 3DUI for virtual environments has to take into account perceptual, cognitive and social issues. Researchers in the social and cognitive field have demonstrated that a wide

CHI 2017, May 6-11, 2017, Denver, CO, USA.

http://dx.doi.org/10.1145/3025453.3025935

range of tasks are better performed in collaborative groups than by individuals alone [13], but there is a lack of this kind of study in the collaborative 3D manipulations field.

One of the main advantages of having a collaborative interface is that it provides users with an individual action perspective of the same scene, which may also include shared and individual viewpoints. In 3D, because of the occlusions, it is impossible to be precise when working from a single perspective (egocentric or allocentric). The accomplishment of simple 3D manipulation tasks, such as a docking, requires constant movement of the user to check occluded parts. This is tedious, especially when the objects' shape is complex. Collaboration between multiple users, in this case, can help in precision and time performances.

In this work, we introduce and assess a novel 3D user interface for collaborative object manipulation. Our technique explores the advantages of smartphones as an interface for 3D interaction. Smartphones are portable, lightweight, wireless and ubiquitous in our contemporaneous life. Most current devices include inertial sensors and a touchscreen, from which one can retrieve the device orientation and multi-touch gestures. The transformation actions we exploit are consolidated in the everyday tasks when interacting with mobile apps and games, such as touch and slide to translate, device orientation to rotate, and pinch and spread to scale the virtual object. Our approach accepts simultaneous connection of multiple smartphones, allowing users to collaborate by simultaneously controlling any manipulation DOFs. This produces equal participation, individual responsibility and positive interdependence, which are necessary features in a collaborative task.

We first assessed the interface with four informal user experiments on docking and obstacle crossing tasks with teams of two users. These experiments were used to refine the interaction, ergonomics and user experience. Then, we conducted a controlled study to understand the influence of group size in collaborative 3D manipulation. We evaluate teams in combinations of one, two, three and four participants. We hypothesize that the team's overall performance increases accordingly to the number of members in the team.

RELATED WORK

Smartphone as 3DUI

The use of smartphones as 3DUI input devices for an external computing environment has already been explored to some extent. Here, we focus on works that use mobile devices to perform 3D manipulations of distant objects displayed on a second screen.

Early work have explored the orientation of a mobile device to intuitively rotate an external object. Katzakis et al. [11] showed a gain in time with the 1:1 mapping of device and virtual object orientation as compared to mouse and touchscreen input. Song et al. [23] used the orientation of a mobile device to rotate a slicing plane in a medical volume data. The mobile device is placed close to a large projection, and the slicing plane extends from the edges of the device. Debarba et al. have explored inertial sensors combined with touchscreens to provide coarse + fine-grained 3D selection [7, 8]. The former reference also includes 2D translation and uniform scaling of distant objects.

Berge et al. [4] have designed a smartphone-based overview+detail interface to interact with 3D public displays. They focus on solutions for the translation of the smartphone detailed view on the overview public screen. They evaluated three interaction techniques to control the movements of the detailed view: classical touchscreen pad, mid-air movements of the mobile device and around the device mid-air hand. Mid-air techniques performed better than touchscreen input, and mid-air phone is most suitable in usual public conditions. In another work [3], Berge et al. extend the mid-air hand into an "Around the smartphone" (ASP) approach to manipulate 3D elements on a second screen. They evaluated the technique against an existing tangible and a tactile implementations in two user studies. Both ASP and tangible performed similarly and better than the tactile technique. Rotation tasks were statistically better performed with the tangible technique. In this technique the rotations follow the wrist movements of the user in a 1:1 mapping. They used a tracking device for rotation calculations instead of the smartphone sensors.

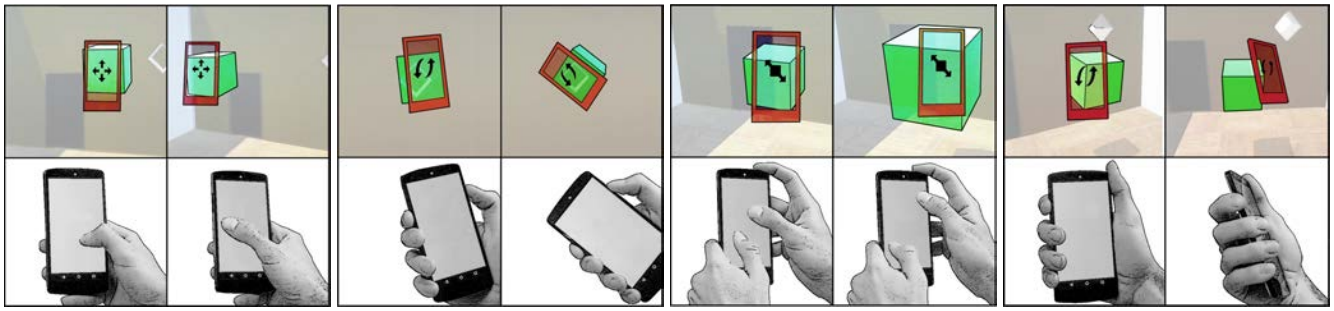
Liang et al. [14] investigated the easiest and most intuitive sensory inputs to use in a tablet, and the most natural interactions to perform in such devices. Although the authors have evaluated the interactions with a tablet, they point out that smaller devices would perform better in motion gestures. A tablet was also the interface for Lopez et al. [15] for their touch-based navigation of 3D visualizations in stereoscopic large displays.

The way we perform the translation transformations is similar to the plane-casting technique presented by Katzakis et al. [12]. Their technique covers rotations and selection as well. They also performed an evaluation compared with an implementation of WAND, classical 6-DOF user interface. They report a significant advantage on translation time using the plane-casting when compared with a wand technique. The smartphone rotation is also explored in Tiltcasting [17], which uses a virtual plane to select 3D objects in crowded virtual environments, and in MobiSweep [26] for creating, modifying, and manipulating 3D shapes in a second screen.

Collaborative Manipulation of 3D Objects in 3DVEs

The demand for spatial collaborative virtual manipulations emerges when single-user tasks become too difficult to perform due to the many degrees of freedom involved and when the task requires the participation of multiple users. It may happen in many application fields, such as simulation and training, and data exploration [20]. A study conducted by Aguerreche et al. [2] compares three main approaches for collaborative 3D objects manipulation in VE: collaborative tangible device, proposed by themselves [1], DOF separation [18] and mean average of the actions [9].

A tangible user interface (TUI) is designed to give a physical form the power to control virtual objects. A TUI can be non-reconfigurable – the physical object cannot be modified, or configurable, where the shape can change. Aguerreche et al. [1] designed a configurable tangible device (CTD) that can be manipulated by single or multiple users to control



(a) Object translation (b) Object rotation (c) Object scale (d) Camera rotation

Figure 2: Walkthrough of manipulations that a mobile phone cursor can perform over the selected object: (a) translation can be applied by touching the mobile phone screen and sliding the point of contact, the object translates on the plane defined by the mobile device orientation; (b) by holding the volume down button and rotating the phone one can rotate the object likewise; (c) scale is applied by touching the screen with two fingers and producing a pinch/spread gesture; (d) the camera orientation can be controlled by holding the volume up button and rotating the mobile phone.

a virtual object in a VE. In a similar manner, Salzmann et al. [21] propose to use a non-reconfigurable tangible device for two-user collaborative interactions. They also compare their technique with a virtual method in an assembly task. Both authors report that a prop-based interaction improves task performance and collaboration because of the link between the two users provided by the tangible device.

DOF separation consists in splitting the tasks among users. In this case, the number of DOFs that each user can access and control is limited: one user controls rotation of the object, while the other one is limited to translation. Pinho et al. [18] explore this approach to demonstrate that the use of cooperative interaction techniques can be more efficient than two users working in parallel using single-user interaction techniques.

The mean technique, in turn, combines user's actions by averaging positions and orientations that they provide. The SkeweR [9] technique enables multiple users to simultaneously grab any part of a virtual object through special points called "crushing points". To determine the translation and the rotation of a grabbed object, SkeweR considers positions of these points and average the 3D object final position. Ruddle et al. [20] propose symmetric and asymmetric solutions to combine two users movements to obtain the virtual object final position.

There are two main differences between previous works and ours. First, we do not restrict the DOFs per participant in the group, but they can organize themselves to split the work. Second, participants' actions are not averaged but summed up in the final object transformation. More details are given in Section *Managing Concurrent Transformations*.

3D MANIPULATION INTERFACE

Graphical Representation

The representation of each mobile phone in the VE includes two elements: (i) a smartphone 3D model that orbits around the selected object following the phone orientation (Figure 2). When the user rotates the device, the virtual cursor (virtual representation of the smartphone) follows, assuming the same orientation and (ii) a picture-in-picture (PIP) camera containing the point of view of the orbiting phone model, positioned at the top right corner of the display (Figure 1). The PIP camera is meant to provide additional depth cues, which are rather limited when the display is shared among multiple users. Each connected device is labeled with a color. The user interface includes a global camera and display, which are shared by all collaborating users. All mobile phones are calibrated with respect to the screen position.

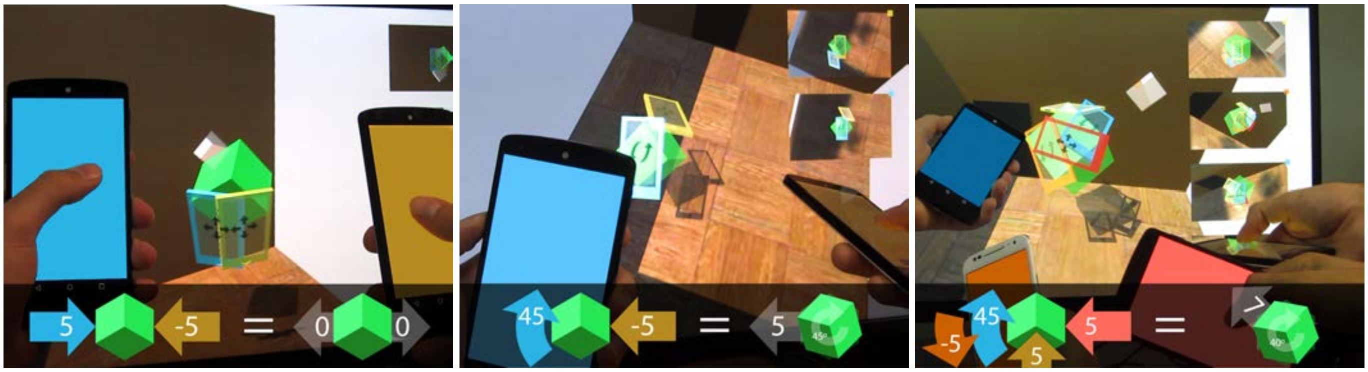
We also render icons on the virtual phone representation indicating the transformation being performed by each user, so that collaborating users can be aware of each other's actions without the need of verbal communication. The icons can indicate that a mobile phone is either performing a translation, rotation or scale on the selected object (Figure 2a-c).

We chose to not use the device display for visualization. It avoids the focus changing between the main screen and the device display that can lead to the losing of context.

Manipulation Mapping

Our 3D manipulation interface uses the mobile phone physical orientation, the touchscreen, and the physical volume buttons to manipulate a total of 7 DOFs of a selected 3D object. More specifically, there are 3 DOFs for translation, 3 DOFs for rotation and 1 DOF for uniform scale.

We use the sensors fusion capability provided by the Android API to retrieve the device orientation. The fusion uses the accelerometer, gyro, and magnetometer to provide an orientation in the 3D space. A calibration action is used once to provide a 1:1 mapping. To translate the selected object, a plane is defined using the device orientation (1:1 mapping). This plane is aligned to the touchscreen and, as the thumb (or any other finger) starts to touch and slide on the screen, a corresponding translation is applied to the object (Figure 2a). The object translation rate is constant and equivalent to 1 unit in the VE for each 333 pixels of sliding over the device's screen. The cube has initial side of 1 unit. The selected object is rotated by pressing and holding the volume down button. Throughout



(a) Forces cancellation

(b) Transformations merge

(c) Concurrent access to the same DOF

Figure 3: Concurrent access to transformations. Every action performed by each user counts as a transformation step. At the bottom of the image we show the applied forces and the resulting transformation.

the time the button is pressed, the virtual object will rotate following the orientation of the phone (Figure 2b). Thus, during a rotation action, the manipulated object follows a 1:1 mapping with the physical device. We do not allow changes in the rotation rate in order to preserve this absolute mapping. Clutch can be usedto reach a total rotation beyond wrist limits (i.e. perform object rotation, reposition the mobile phone, then perform a new object rotation). Uniform scale is performed by pinch and spread gestures with two fingers on the touchscreen (Figure 2c). The scale is obtained by dividing the current distance between the fingers by their initial distance. Finally, the main camera (shared among users) follows the selected object at a fixed distance. If a user wishes to change the camera orientation, this can be done using the volume up button in a similar manner as the virtual object rotation action (Figure 2d). The virtual shared camera can be only rotated in the pitch and yaw axes. If users wish a personalized point of view, they can use the embedded PIP camera.

A user can perform any transformation action isolated or two by two: rotation and translation, scale and rotation, camera rotation and translation, scale and camera rotation.

GENERALIZING FOR COLLABORATIVE 3D MANIPULATION

Managing Concurrent Transformations

We do not impose limits on the number of simultaneous users. All users have access to all available functions. While working in groups, collaborating users may wish to split the different aspects of the manipulation among them. For instance, while the first user translates an object, the second user could rotate this same object.

All user actions are stored in a matrix representation and sent through the network to the server. Every action performed by each individual user counts as a transformation step. Action matrices are multiplied by the transformation matrix of the virtual object. Therefore, every contribution from each user is summed up into the final object's transformation without restrictions or weights. Therefore, if two users move the object in opposite directions, the position of the object will not change (Figure 3a). On the other hand, if they manipulate the different transformations in parallel, the transformations are merged (Figure 3b). Figure 3 shows the concurrent access of same transformations by three users.

Since all the users' actions are applied directly to the manipulated virtual object, we smooth the movements with an exponential moving average filter [22] to minimize the undesired flickering. It was chosen because it only needs the previous values to filter the signal. In this work, the filter is in the format: $NewValue = PrevValue * (1 - C) + CurrentSignal * C$. Where C is the filter constant. We fixed $C = 0.5$.

COLLABORATIVE 3D MANIPULATION ASSESSMENT

We conducted four public demonstrations of our collaborative user interface prior to the user study presented below. In these demonstrations we proposed different 3D manipulation tasks: object docking, and obstacle crossing. In the former, the users had to manipulate 7-DOF of selected objects in order to stack them by precisely docking one on top of the other. In the latter, the users had to manipulate a selected cube to take it from an initial to a destination position. The cube had to be carried through a sequence of walls with openings while avoiding collisions but occupying the maximum volume possible. Surface constraints were used to prevent the interpenetration between the cube and the obstacles. Over a hundred people have tested our technique in the demonstrations. We collected users' feedback to refine the interaction, ergonomics and user experience.

The obstacle crossing task was then used for a formal experiment to assess the effect of group size on manipulation time and accuracy. Three tasks are part of the obstacle crossing experiment. The first task is a wall with a square opening scaled down but aligned to the axes of the manipulated cube. The second is a wall with a both scaled and rotated square opening. The third is a winding tunnel with cross sections that vary in width and orientation (Figure 4). Each task starts after a three seconds count down and finishes when the object reaches the destination area at the end of the obstacle. The first two walls are used for training and the tunnel is used for the evaluation.

The evaluation consists in calculating the time and the accuracy to transpose the tunnel. We calculated the time from the start to the end of the task and the accuracy through a checkpoint system. The checkpoints are objects identical to the manipulated object (yellow cubes in Figure 4). We placed eight checkpoints along the tunnel path with a static position, orientation and scale so as to occupy the maximum volume at that location. To perform the task with high accuracy the users had to occupy the maximum volume of these checkpoints with the manipulated cube while completing the circuit. To calculate the error between the manipulated object m and a checkpoint c, we take the minimum distance from a vertex m_i to all vertexes in c. Thus, the error for each checkpoint is the maximum distance among the minimum distances previously calculated. For each task, we compute the error median for the eight checkpoints.

To the best of our knowledge, there is no well established interface for 3D collaborative manipulation (translation, rotation and scaling) to permit a comparison with our technique. We are aware of slightly distinct input-output mappings of mobile input for some of these transformations in a single user mode, though, from very good results regarding precision and comfort [7, 8]. We used these previous results to justify our design choices and decided to focus on the evaluation of the collaborative aspects of our technique.

Figure 4: Checkpoints in the obstacle crossing task.

USER STUDY

Design and Procedure

We aim to investigate the relationship between group sizes and the time and accuracy to complete the tasks. Furthermore, we intend to understand the influence of work distribution balance and work division in the performance of each group combination. Thus, the experiment follows a between subject design with *Group size* as the only independent variable, with one, two, three or four participants. Dependent variables collected were *time* to complete the task and *accuracy* of the group, and *transformation actions* (translation, rotation, scale or camera rotation), including duration and magnitude of the action performed by each individual subject. The accuracy is measured as described before in Section *Collaborative 3D Manipulation Assessment*.

Participants answered a characterization form before the experiment. They also watched an informative 2 minutes video about the technique and the task. Then, we handed the mobile devices already connected in the virtual environment to the participants. We demonstrated the calibration steps and, after that, we started with the training session. The training session began with an individual exercise and continued with the group practice. The session finished when all participants were satisfied, without time restrictions. Then, users recalibrated the devices (when needed) and we started the test. Between the trials, users could recalibrate the devices as well. We asked participants to prioritize accuracy over time. The test ends with a post-experiment questionnaire. In average, the experiment sessions lasted 30 minutes.

Our first hypotheses are:

- H1. Groups with more than one member complete the tasks faster
- H2. Groups with more than one member complete the tasks with more accuracy
- H3. For the tested group size range, if groups increase in members, the time to complete tasks drops proportionally
- H4. For the tested group size range, if groups increase in members, the accuracy to complete tasks increase proportionally

If the hypotheses are confirmed, we can assume that our interface provides a significant gain in performance through collaborative work.

Task

We used the obstacle crossing game with three wall configurations, as described in Section *Collaborative 3D Manipulation Assessment*. The training sessions consist of the first two walls. The test session is formed by one trial for each practice wall and two trials for the tunnel. In all trials, the selected object starts 4 units distant from the walls. After the obstacle, the object has to be carried to a finishing wall distant 4 units.

The Section *Results* reports only the two trials in the tunnel task for the statistical analysis.

Subjects

Sixty subjects participated voluntarily in this experiment (nine female), aged 24 years in average (SD=3.6). They were all Computer Science students with no movement restrictions on wrists and arms. Thirteen of the individuals had never used gestural interactions with Kinect, Wiimote or mobile devices. We arranged the participants in 5 groups of one, 7 groups of two, 7 groups of three and 5 groups of four individuals.

Experimental Setup

Our setup is composed of mobile devices, a server computer, a projection screen and a WiFi router. An app on the phone communicates with the server that manages the client's data and the virtual environment. In the experimental setup, the same screen was shared by all the users.

Although we used diverse handheld devices, all of them were Android-based smartphones with the 5.0+ version of the operating system and had WiFi connection. They had physical volume buttons placed on the side of the device, touch screen, and gyroscope and accelerometer sensors. The same compilation of the application was installed on all devices.

The server application runs on a PC and is developed on Unity 3D. It renders the virtual environment, manages the connections with the mobile phones, and the manipulation of the transformations to be performed. Communication between server and phones is via a dedicated WiFi router and uses TCP network protocol to ensure that the packets will arrive in the correct order.

The projector is a Sharp with 1024x768 resolution. It is placed at 225 cm above the participants and 260 cm far from the projection screen. The projected image has 143x110 cm in size. The participants are placed next to each other at a distance of 220 cm from the screen. Figure 5 shows the physical space setup.

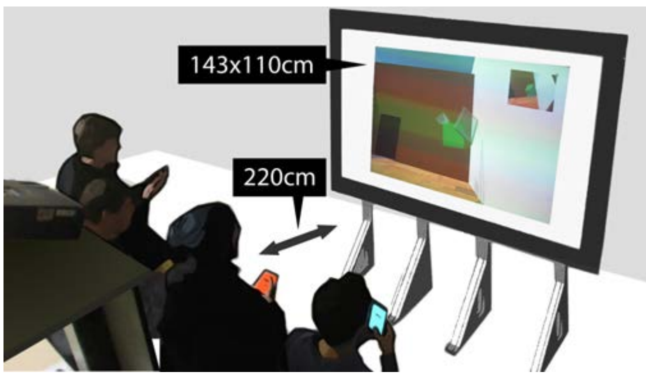


Figure 5: Physical setup of the experiments. Participants are placed next to each other at a distance of 220 cm from the screen. Each of them has its own device.

RESULTS

Group Size vs. Time and Accuracy

Figure 6 shows the time and accuracy that each group configuration achieved in the evaluation. For the statistical analysis, we first verified if the relation between the independent (group size) and the dependent variables (time and accuracy) could be evaluated using ANOVA. We fitted a linear regression and tested the normality of the residuals (Shapiro-Wilk test). Residuals were not normally distributed on both comparisons. Although Norman [16] suggests that ANOVA is robust when residuals are not normally distributed, we made a conservative choice by using non-parametric tests as these make claims about the difference of medians instead of the average.

We conducted a Kruskal-Wallis test, which can determine if a statistically significant difference exists between the levels of *group size* without assuming residuals to follow the normal distribution. Post-hoc was conducted using paired Dunn tests with Holm-Bonferroni correction.

Group Size vs. Task Completion Time

The Kruskal-Wallis test failed to reject equality of medians across different group sizes for task completion time ($H(3) = 2.1834$, $p = 0.54$), thus we reject H1 and H3. See Figure 6a.

Group Size vs. Task Accuracy

As we hypothesized in H2, the Kruskal-Wallis test revealed significant effect of group size on task accuracy median ($H(3) = 21.3522$, $p < 0.0001$). The post-hoc Dunn test indicates that significant accuracy increase occurs between group sizes 1 and 3 ($p = 0.0486$), 1 and 4 ($p < 0.0001$), 2 and 4 ($p = 0.0024$) and 3 and 4 ($p = 0.0226$). This result confirms H4. No significant difference was measured between group sizes 1 and 2, and between group sizes 2 and 3. See Figure 6b.

Groups vs. Work Division

Accuracy vs. Work Distribution Balance

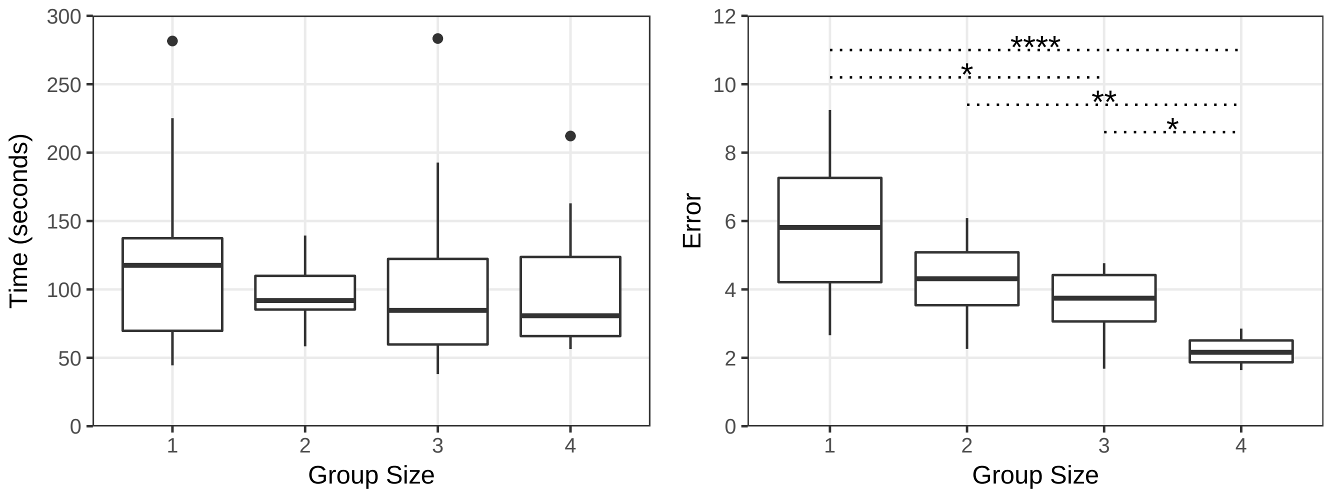
For a more complete understanding on how the various group sizes affect accuracy, we analyzed the balance of the work distribution among members of teams with different sizes. Our hypothesis is that the balance of work distribution affects the task accuracy. The work distribution balance is a value between zero and one, where zero is the minimum work distribution and one is the maximum work distribution.

Before the work distribution evaluation, we performed an analysis to assess the workload of each team members for all group sizes. Each participant's workload was quantitatively measured dividing the user's active time by the group's active time. The Kruskal-Wallis test revealed a significant effect in the workload when group sizes vary ($H(3) = 79.0784$, $p < 0.0001$). The Dunn post-hoc indicates significant decrease in workload between groups size 1 and 2 ($p < 0.0006$), 1 and 3 ($p < 0.0001$), 1 and 4 ($p = 0.0031$), 2 and 3 ($p < 0.0007$), 2 and 4 ($p < 0.0001$) and 3 and 4 ($p = 0.0303$).

Then, we computed the work distribution balance. We calculated the group variance using each individual workload previously calculated. We adjusted the results by multiplying the variance by the respective group size. In this test, only groups with two, three and four members were evaluated, since the work distribution in groups with one participant is zero. However, the Kruskal-Wallis test failed to reject equality of medians across different group sizes for work distribution balance ($H(2) = 4.482$, $p = 0.11$). This hinders our intent of correlating accuracy with overall work distribution balance directly. Nevertheless, this result permitted us to focus on action division patterns, as detailed next.

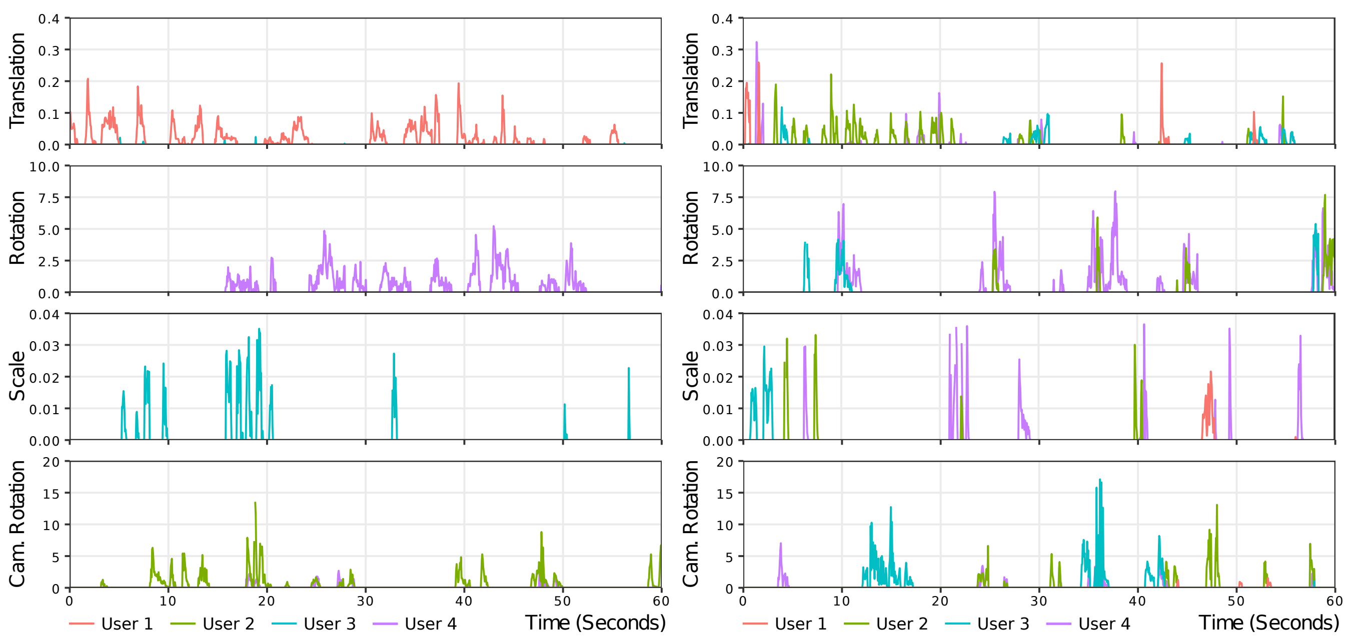
Groups vs. User Roles

The work distribution balance omits the group action division by calculating a single distribution score for the whole group. Here, we analyze each team member individually to find a work division pattern. Our hypothesis is that the groups that divide actions among members are more accurate than the groups that do not divide. The time series on Figure 7 shows the different strategies adopted by two groups with four participants, each with similar work division balance (A = 96.4%, B = 95.5%) but with different action strategies and accuracy. To



(a) (b)

Figure 6: (a) There was no significant difference in time to complete the tasks between groups. (b) Group size 4 is the most accurate (Mdn = 2.16, SD = 0.44), followed by size 3 (Mdn = 3.74, SD = 0.94), size 2 (Mdn = 4.31, SD = 1.18), and size 1 (Mdn = 5.81, SD = 2.16).



(a) Team A (b) Team B

Figure 7: Time series of one task showing the collaboration strategy adopted by Team A and B with four participants. Both teams have similar work division balance (A = 96.4%, B = 95.5%) but different strategies. Team A splitted the transformations among the participants in such a way that, on average, all team members performed 3.25 roles swap. They completed the task in 60.7 seconds and with 1.85 errors. Team B adopted a division strategy, on average all team members performed 15 roles swap. They completed the task in 63.9 seconds (the last 3.9s are omitted in the plot) and with 2.78 errors.

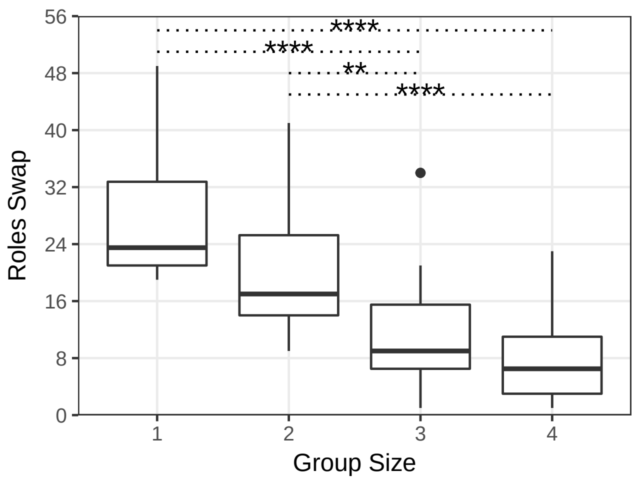


Figure 8: Individuals in groups with four members swap roles less often (Mdn = 6.5, SD = 6.26), followed by groups with size 3 (Mdn = 9, SD = 7.21), size 2 (Mdn = 17, SD = 10.22) and size 1 (Mdn = 23.5, SD = 10.71).

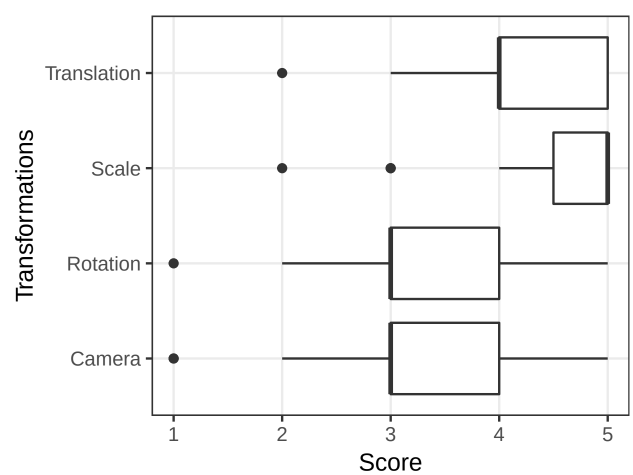


Figure 9: User's opinion about the technique comfort for each transformation. The results are reported in the Likert scale where 1 is very uncomfortable and 5 is very comfortable. Translation (Mdn = 4, SD = 0.73), scale (Mdn = 5, SD = 0.65), rotation (Mdn = 3, SD = 0.94) and camera rotation (Mdn = 3, SD = 1.08).

extract the division of actions between the groups, we identified the frequency users change between actions (translation, rotation, scale). We call this change a *role swap*.

The Kruskal-Wallis variance analysis revealed significant effect between groups and user roles changes ($H(3) = 40.1615$, $p < 0.0001$). The post-hoc Dunn test indicates that significant action division occurs between groups size 1 and 3 ($p < 0.0001$), 1 and 4 ($p < 0.0001$), 2 and 3 ($p = 0.0031$) and 2 and 4 ($p < 0.0001$). Groups size 1 and 2, 3 and 4 do not differ significantly (see Figure 8).

Accuracy vs. User Roles

Since the accuracy and the action division have similar behaviors, we hypothesized that the two variables were related. The Pearson correlation revealed a significant effect between error and user roles changes ($r = -0.3957$, $p < 0.0001$).

Learning Between Trials

In Figure 10, we measured the learning effect between the two trials. We assessed the time and accuracy for all groups together. The Wilcoxon signed rank test indicated no significant increase in accuracy between trials ($Z = 144$, $p < 0.5879$) and a significant effect in time between trials ($Z = 214$, $p = 0.0032$).

User Comfort

In Figure 9, we report the user responses to comfort when performing the transformations. Most users reported that the translation and scale actions are very comfortable to perform. Object and camera rotations had mostly neutral or positive opinions. They have almost the same comfort level according to users, being slightly but clearly below the other two transformations.

DISCUSSION

Technique Performance

The findings in our study indicate that handheld devices are powerful tools for 3D collaborative tasks. We observed that team arrangements with two, three and four members solve the tasks significantly more accurately than individuals (groups with one member). Interestingly, the accuracy increases faster with larger group sizes. For group sizes of one and two, there is only significant increase when compared with groups with two or more additional members. However, between group sizes 3 and 4, the accuracy increases significantly. No significant difference in completion time is reported, which may have been influenced by the instruction to prioritize accuracy over time in our experiment.

To understand the causes for accuracy increasing with larger group sizes, we investigated whether it was caused by the way teams balance the actions among team members. The statistical analysis revealed no significant increase. Thus, we further explored the accuracy effect by analyzing the team's work division. We observed a significant drop in the swap of roles of the team members on larger groups. It indicates that users tend to specialize in one transformation, consequently better dividing the tasks. The behavior between accuracy and work division follows the same trend, as observed in Figure 6b and Figure 8. We correlated the two variables, and the results indicate that the work division strategy is related to the error drop.

We also investigated the learning effect between the two repeated trials. We report a significant drop in time to complete the tasks without affecting accuracy. It indicates that with more training, groups tend to perform the task faster while keeping equivalent accuracy.

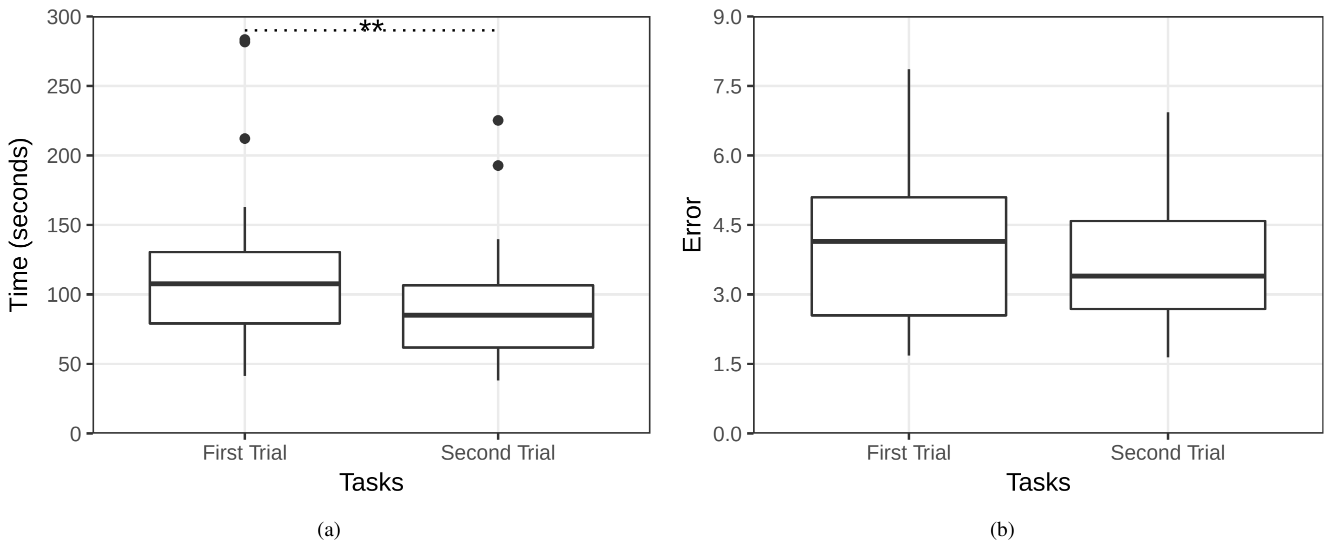


Figure 10: (a) Groups complete the task significantly faster in the second trial (Mdn = 85.12, SD = 44.81) than the first (Mdn = 107.59 , SD = 65.58). (b) There is no significant accuracy effect between the first (Mdn = 4.14, SD = 1.74) and the second (Mdn = 3.59, SD = 1.83) trials.

Technique Comfort

The results shows that the technique is comfortable. We observed that translation and scale receive better scores than both object and camera rotations. It is known that actions performed by small muscle groups, such as fingers tend to be more comfortable than larger muscle groups, such as wrist and arm [28]. This may explain why the rotation actions were ranked one step below the others, even if still well ranked. Nevertheless, we exchanged comfort for affordance in our design. We believe that the use of direct mapping for rotations is more natural than rotations with the fingers, as the user indirectly holds the virtual object.

Ringelmann Effect

The Ringelmann Effect states that the addition of new coworkers in a collaborative task (the original was a rope pulling task) leads to a linear decrement in the member's performance. Ingham et al. [10] reproduced the effect with teams of 1 to 6 participants. They report a significant drop in individual performance in groups with more members. Even though our experiment was not designed to test the effect, in Section *Accuracy vs. Work Distribution Balance*, we found a correlation between workload and group size. It suggests that, in virtual collaboration, team members tend to work less in larger groups as previously demonstrated in the real world collaborative tasks.

CONCLUSION

In this paper, we presented the design of a novel 3D user interface for collaborative object manipulation in 3D virtual environments. The technique is based on smartphones and uses the touchscreen and the inertial sensors as a 3DUI. The technique has proven to be intuitive and robust. More than a hundred participants have tested it and could provide data and feedback about its usability. In several informal tests, users were invited to download the app in their own smartphones and to join an on-going manipulation session. We tested it with teams composed of more than ten members using smartphones of different models. All of them easily and quickly understood the purpose and mechanics of the technique, indicating a high affordance.

Smartphones, being versatile and ubiquitous devices, conveniently adapt as input devices to be used in combination with an external or wearable display. While we evaluated our interaction technique using a large screen (integrating a shared+personal view) as display device, the technique is friendly to other display form factors such as HMDs or CAVEs. Changing the view to VR HMDs or augmented reality HMDs (e.g. Hololens) could provide greater individual perspective, while the input manipulations would not change much.

In formal user experiments, we demonstrated that the technique design is suitable to be used by teams. Teams with more members performed more accurately than smaller groups. In larger groups, members could better divide the basic tasks, assuming roles and, as a consequence, decreasing the workload. Speed was not the goal in our experiments, but we also observed that speed increases from the first use of the system to the second one without affecting accuracy, demonstrating that the users learn fast. However, more tests should be done regarding the technique learning curve.

In this study, we aimed at understanding how groups behave in a situation where users have freedom to perform any transformation at any time. In a future work, we plan to focus on different levels of controlled labor division in a way that each user will be responsible for only one specific type of transformation. We expect that, by forcing the role division between the peers, the accuracy could be increased for all the groups.

ACKNOWLEDGMENTS

We thank all the users that volunteered for the experiment. Thanks are also due to Iago Berndt for his help with the implementation and Victor Adriel de Jesus Oliveira for his help with the text and analysis revision. We finally acknowledge the funding from CAPES, CNPq (305071/2012-2), and FAPERGS (2283-2551/14-8).

REFERENCES

1. Laurent Aguerreche, Thierry Duval, and Anatole Lécuyer. 2010. Reconfigurable Tangible Devices for 3D Virtual Object Manipulation by Single or Multiple Users. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology (VRST '10)*. ACM, New York, NY, USA, 227–230. DOI: http://dx.doi.org/10.1145/1889863.1889913
2. Laurent Aguerreche, Thierry Duval, and Anatole Lécuyer. 2011. Evaluation of a reconfigurable tangible device for collaborative manipulation of objects in virtual reality. In *Theory and Practice of Computer Graphics*. 81–88.
3. Louis-pierre Bergé, Emmanuel Dubois, and Mathieu Raynal. 2015. Design and Evaluation of an "Around the SmartPhone" Technique for 3D Manipulations on Distant Display. In *SUI15*. 69–78. DOI: http://dx.doi.org/10.1145/2788940.2788941
4. Louis-Pierre Bergé, Marcos Serrano, Gary Perelman, and Emmanuel Dubois. 2014. Exploring Smartphone - Based Interaction with Overview + Detail Interfaces on 3D Public Displays. *MobileHCI 2014* (2014), 125–134. DOI: http://dx.doi.org/10.1145/2628363.2628374
5. Doug A Bowman, Ernst Kruijff, Joseph J LaViola, and Ivan Poupyrev. 2004. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA.
6. Elizabeth F Churchill and Dave Snowdon. 1998. Collaborative virtual environments: an introductory review of issues and systems. *Virtual Reality* 3, 1 (1998), 3–15.
7. Henrique Debarba, Luciana Nedel, and Anderson MacIel. 2012. LOP-cursor: Fast and precise interaction with tiled displays using one hand and levels of precision. *IEEE Symposium on 3D User Interfaces 2012, 3DUI 2012 - Proceedings* (2012), 125–132. DOI: http://dx.doi.org/10.1109/3DUI.2012.6184196
8. Henrique G. Debarba, Jerônimo G. Grandi, Anderson Maciel, Luciana Nedel, and Ronan Boulic. 2013. Disambiguation Canvas: A Precise Selection Technique for Virtual Environments. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 8119 LNCS. 388–405. DOI: http://dx.doi.org/10.1007/978-3-642-40477-1_24
9. Thierry Duval, Anatole Lécuyer, and Sébastien Thomas. 2006. SkeweR: A 3D interaction technique for 2-user collaborative manipulation of objects in virtual environments. In *3DUI 2006: IEEE Symposium on 3D User Interfaces 2006 - Proceedings*, Vol. 2006. 69–72. DOI: http://dx.doi.org/10.1109/VR.2006.119
10. Alan G Ingham, George Levinger, James Graves, and Vaughn Peckham. 1974. The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology* 10, 4 (1974), 371–384.
11. Nicholas Katzakis and Masahiro Hori. 2009. Mobile phones as 3-dof controllers: A comparative study. In *Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on*. IEEE, 345–349. DOI: http://dx.doi.org/10.1109/DASC.2009.76
12. Nicholas Katzakis, Robert J Teather, Kiyoshi Kiyokawa, and Haruo Takemura. 2015. INSPECT: extending plane-casting for 6-DOF control. *Human-centric Computing and Information Sciences* 5, 1 (dec 2015), 22. DOI: http://dx.doi.org/10.1186/s13673-015-0037-y
13. Patrick R. Laughlin. 2011. *Group Problem Solving*. Princeton University Press.
14. Hai Ning Liang, Cary Williams, Myron Semegen, Wolfgang Stuerzlinger, and Pourang Irani. 2013. An investigation of suitable interactions for 3D manipulation of distant objects through a mobile device. *International Journal of Innovative Computing, Information and Control* 9, 12 (2013), 4737–4752.
15. David Lopez, Lora Oehlberg, Candemir Doger, and Tobias Isenberg. 2016. Towards An Understanding of Mobile Touch Navigation in a Stereoscopic Viewing Environment for 3D Data Exploration. *IEEE Transactions on Visualization and Computer Graphics* 22, 5 (may 2016), 1616–1629. DOI: http://dx.doi.org/10.1109/TVCG.2015.2440233
16. Geoff Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education* 15, 5 (2010), 625–632.
17. Krzysztof Pietroszek, James R. Wallace, and Edward Lank. 2015. Tiltcasting: 3D interaction on large displays using a mobile device. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology - UIST '15* (2015), 57–62. DOI: http://dx.doi.org/10.1145/2807442.2807471
18. Marcio S Pinho, Doug A Bowman, and Carla M Dal Sasso Freitas. 2008. Cooperative object manipulation in collaborative virtual environments. *Journal of the Brazilian Computer Society* 14 (2008), 54–67. http://www.scielo.br/scielo.php?script=sci
19. Mike Robinson. 1993. Designing for Unanticipated Use.... In *ECSCW '93*. 187–202. DOI: http://dx.doi.org/10.1145/958217.958218
20. Roy A Ruddle, Justin C D Savage, and Dylan M Jones. 2002. Symmetric and asymmetric action integration during cooperative object manipulation in virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)* 9, 4 (2002), 285–308.

21. H. Salzmann, J. Jacobs, and B. Froehlich. 2009. Collaborative interaction in co-located two-user scenarios.. In *Proc. of JVRC 2009 (Joint Virtual Reality Conference of EGVE - ICAT - EuroVR)*. 85–92.

22. Steven Smith. 2013. *Digital signal processing: a practical guide for engineers and scientists*. Newnes.

23. Peng Song, Wooi Boon Goh, Chi-Wing Fu, Qiang Meng, and Pheng-ann Heng. 2011. WYSIWYF: Exploring and Annotating Volume Data with a Tangible Handheld Device. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, Vol. D. ACM Press, New York, New York, USA, 1333. DOI: http://dx.doi.org/10.1145/1978942.1979140

24. Wolfgang Stuerzlinger and Chadwick A. Wingrave. 2011. The value of constraints for 3d user interfaces. *Virtual Realities: Dagstuhl Seminar 2008* (2011), 203–223. DOI: http://dx.doi.org/10.1007/978-3-211-99178-7_11

25. Manuel Veit, Antonio Capobianco, and Dominique Bechmann. 2009. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. *Proceeding VRST '09 Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* 1, 212 (2009), 51–58. DOI: http://dx.doi.org/10.1145/1643928.1643942

26. Vinayak, Devarajan Ramanujan, Cecil Piya, and Karthik Ramani. 2016. MobiSweep: Exploring Spatial Design Ideation Using a Smartphone as a Hand-held Reference Plane. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '16*. ACM Press, New York, New York, USA, 12–20. DOI: http://dx.doi.org/10.1145/2839462.2839490

27. Shumin Zhai and Paul Milgram. 1998. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '98* (1998), 320–327. DOI: http://dx.doi.org/10.1145/274644.274689

28. Shumin Zhai, Paul Milgram, and William Buxton. 1996. The influence of muscle groups on performance of multiple degree-of-freedom input. *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96* (1996), 308–315. DOI: http://dx.doi.org/10.1145/238386.238534