The Effect of DOF Separation in 3D Manipulation Tasks with Multi-touch Displays

Anthony Martinet* Géry Casiez † Laurent Grisoni ‡
LIFL & INRIA Lille
University of Lille, FRANCE

Abstract

Multi-touch displays represent a promising technology for the display and manipulation of data. While the manipulation of 2D data has been widely explored, 3D manipulation with multi-touch displays remains largely uncovered. Based on an analysis of the integration and separation of degrees of freedom, we propose a taxonomy for 3D manipulation techniques with multi-touch displays. Using that taxonomy, we introduce *DS3* (*Depth-Separated Screen Space*), a new 3D manipulation technique based on the separation of translation and rotation. In a controlled experiment, we compare *DS3* with *Sticky Tools* and *Screen-Space*. Results show that separating the control of translation and rotation significantly affects performance for 3D manipulation, with *DS3* being at least 22% faster.

CR Categories: H.5.2 [User Interfaces]: Input devices and strategies.

Keywords: Multi-touch displays, 3D manipulation task, direct manipulation

1 Introduction

Three-dimensional (3D) manipulation is an important challenge for 3D user interface designers, involving the control of six Degrees Of Freedom (DOF) : three for position (i.e. translation along x, y and z-axis) and three for orientation (i.e. rotation around x, y and z-axis). Using current desktop interfaces and a mouse, rotating 3D objects can take from ten to thirty seconds [Hinckley et al. 1997], much slower than real object manipulation which takes between one and two seconds [Wang et al. 1998].

Compared to the mouse, multi-touch displays provide extra input bandwidth through multiple contact points and also enables direct manipulation by allowing users to directly *touch* data [Rekimoto 2002]. The Rotate-Scale-Translation gesture (RST) for manipulating 2D data is a typical example of such an interaction paradigm [Hancock et al. 2006]. While 2D manipulation on multi-touch displays has been widely explored, 3D manipulation remains largely uncovered. This may be explained by the difficulty of mapping input contact points to the attributes controlled in the 3D task because of the inherent 2D nature of the input device.

*e-mail: anthony.martinet@lifl.fr
†e-mail: gery.casiez@lifl.fr
‡e-mail: laurent.grisoni@lifl.fr

VRST 2010, Hong Kong, November 22 – 24, 2010.
© 2010 ACM 978-1-4503-0441-2/10/0010 $10.00

Jacob et al. [1994] studied the impact on performance of the input device control structure and the perceptual structure of the task. The input device control structure and the perceptual structure of the task can be integral or separable. They found a strong relationship between the two structures with better performance when both match. While it has been shown that human fingers have separable DOF [Ingram et al. 2008], 3D manipulation is inherently an integral task [Garner 1974]. The thumb, index and middle fingers can be moved separately from one another while users perceive the attributes of 3D objects (position and orientation) as a whole. This mis-match between the separable input structure of multi-touch devices and integral nature of a 3D manipulation task raises the question of optimizing the mapping between the two structures.

While techniques like *Sticky Tools* propose a way to separate the DOF of the 3D manipulation task [Hancock et al. 2009], other techniques like *Screen-Space* present a method to integrate them [Reisman et al. 2009]. However the lack of user-study makes it difficult to compare the two approaches.

After presenting the related work on degrees of freedom integration and separation, as well as 3D manipulation techniques with multi-touch displays, we will introduce a taxonomy to compare 3D manipulation techniques with multi-touch displays. We then introduce a new technique called *DS3* (*Depth-Separated Screen Space*) based on the clear separation between translation and rotation. We present the results of a controlled experiment comparing this technique with *Sticky Tools* and *Screen-Space*. Finally the question of controlling integrated DOF with a separable multi-touch input device is addressed in the discussion.

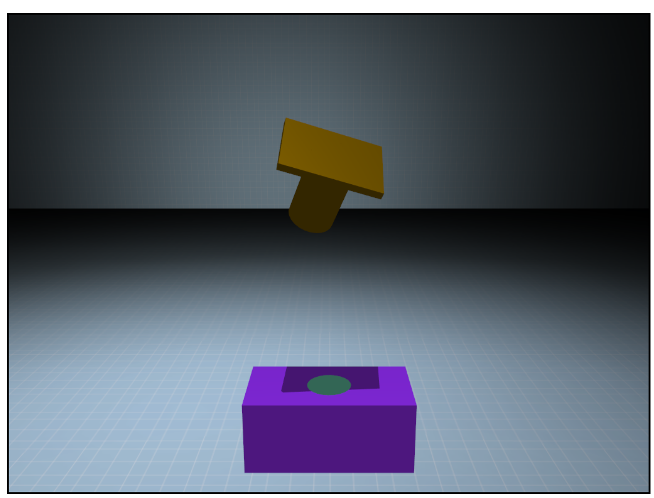

Figure 1: *Screen capture of the peg-in-hole task.*

2 Related work

2.1 Integration and Separation of DOF

According to the theory of perceptual structure of visual information by Garner [1974], a multi-dimensional object can be characterized by its attributes in two categories : integral structure and separable structure. Visual information has an integral structure if its attributes can be perceptually combined to form a unitary whole. If visual object attributes show perceptually distinct and identifiable dimensions, they are separable. According to this definition, the orientation and the position of a 3D object are two integral attributes, making 3D manipulation an integral task.

Jacob et al. [1994] extended Garner's notion of integral and separable structure to interactive tasks by observing that manipulating a graphic object is simply the modification of the values of its attributes. They also extended the notion of integral and separable structure to describe the attributes of an input device, based on whether it is natural to move diagonally across all dimensions. With an integral device, the movement is in Euclidean space and cuts across all the dimensions of control. A separable device constrains movement along one dimension at a time. They conducted an experiment in which subjects performed two tasks that had different perceptual structures, using two input devices with correspondingly different control structures, an integral three-dimensional tracker and a separable mouse. Their results support their hypothesis: human performance increases when the perceptual structure of the task matches the control structure of the device. They concluded that the interplay between task and device was more important in determining performance than either task or device alone.

Wang et al. [1998] extended this theory to extrinsic (i.e. orientation and position) properties of object being manipulated by the human hand. They pointed out that the visual pathways responsible for object perception are separated from those guiding the action. They ran an experiment that asked participants to dock a cube using different visual feedback conditions. They reported that users had little difficulty in simultaneous control of object translation and orientation.

Considering an orientation task only, Veit et al. [2009] studied the integration of DOF. They conducted an experiment in which users had to orient 3D objects with two interaction techniques, one integrating and the other separating the DOF of the orientation task. The results suggest that the simultaneous manipulation of all the DOF does not necessary lead to the best performance, leading to conclusions at the opposite of Jacob et al's.

Regarding 2D manipulation with multi-touch displays, Nacenta et al. [2009] addressed a common problem concerning integration of DOF. When using multi-touch gestures, performing a subset of available operations may be difficult for users. For example, it can be hard to only scale and translate an object (without rotating it) because the object will also react to small variations of the angle between the contact points. They introduced an interaction technique that allows users to select a subset of manipulations for 2D data, reducing unwanted manipulation without negatively affecting performance. Separating the control of DOF leads this time to improve the user's expectations.

2.2 3D Manipulation and Multi-touch Displays

Limited 3D manipulation

Using a standard vision-based tabletop, Wilson et al. [2008] presented a physics-enabled 3D environment with multi-touch input manipulating the DOF of the task in an integral way. Their technique is able to model both multiple contact points and more sophisticated shape information, such as the entire hand. They showed that their technique can be used to add real-world dynamics to interactive surfaces. While the underlying physical simulation can provide a number of convincing effects during the interaction (inertia, collision), integrating all the DOF in such an environment prevents users from lifting object (i.e. move the object along the z-axis).

Hilliges et al. [2009] worked with depth-sensing camera to solve this issue. They provided users with the ability to "pick-up" an object and manipulate it above the surface. This supports Jacob's conclusions since they use extra input information to improve the matching between the control structure and the task perceived structure, both integral in this case. However, those techniques need additional hardware making the compatibility with existing multi-touch displays difficult.

Hancock et al. [2007] presented one, two and three touch input interaction techniques to manipulate 3D objects on any multi-touch display. With three-touch interaction, users can perform simultaneous translation and rotation on the surface of the table. Depth-positioning is proposed as an option by measuring the distance between two fingers. The three-touch technique, called *Shallow-Depth*, was shown to be faster and more accurate as well as preferred by the users. Nevertheless, the 3D task performed required the user to manipulate only 5 DOF.

Martinet et al. [2010] proposed two techniques for 3D positioning. One technique, the Z-technique, presented 3D data in full-screen while the other technique split the screen in 4 viewports. They conducted a docking task experiment but were not able to draw conclusions on performance. As a qualitative point a view, they reported that users preferred the full-screen technique.

Full 3D manipulation

To control all the DOF required for 3D manipulation, Hancock et al. [2009] introduced a new technique called *Sticky Tools*, allowing users to manipulate an object using three fingers. Each finger separately controls DOF that are integrated together. While the authors discussed the use of such a technique in a more general manner, the choice of DOF to integrate together is not addressed and no user study was carried out to measure the efficiency of *Sticky Tools*.

Reisman et al. [2009] introduced a method to handle 3D manipulation in a direct way, integrating all the DOF needed to perform such an operation. Highlighting the fact that RST has become the *de facto* standard technique to handle 2D objects, they presented a technique to extend RST into 3D. The tool consists in solving constraints fixed by users' fingers. A constraint solver minimizes the error between contact points' screen-space projection (i.e. fingers' position on the 2D screen) and their corresponding screen-space target positions (i.e. the 3D points *touched* by fingers). The paper discusses the use of the constraint solver and gives examples to use this tool to design interaction techniques — but no formal evaluation was performed.

3 A taxonomy of 3D Manipulation Techniques with Multi-touch Displays

The 3D manipulation techniques mentioned above control different sub-sets of DOF depending on the number of fingers in contact with the surface. In addition a finger can be considered either *direct* or *indirect* depending on the euclidian physical distance between the finger position and the projection on screen of the virtual object

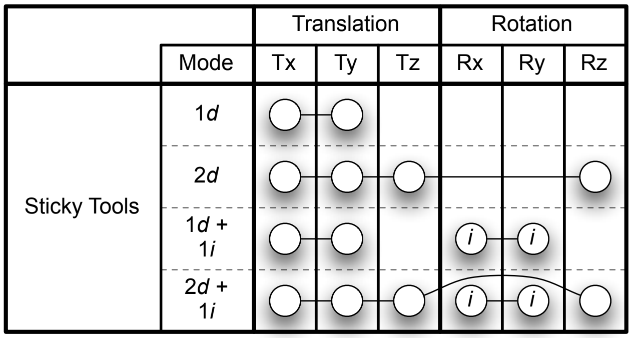

Figure 2: *Description of the Sticky Tools technique using the taxonomy.*

being manipulated. When this distance is equal or close to zero, the finger is *direct* and turns *indirect* when this distance becomes greater. The number of fingers used for the interaction and the directness of each finger is below referenced as a mode.

To help comparing existing manipulation techniques we chose to adapt the taxonomy introduced by Card et al. [1991]. We wanted to represent the relationship between the number of fingers, their directness (whether *direct* or *indirect*) and the corresponding DOF controlled in the task. We also wanted to represent whether the DOF of the task are controlled in an integral or separable way. The degrees of freedom controlled in the manipulation task are represented in a cartesian direct framework where the *x* axis belongs to the screen plane and is oriented towards the right and the z axis is orthogonal to the screen and points towards the user. T_x, T_y and T_z represent the translations along the corresponding axis; R_x, R_y and R_z the rotations around the corresponding axis. This taxonomy only takes into account the information available through the inputs provided by the current technology: the number of contact points (e.g. the number of fingers) and the directness of each finger. This taxonomy could be enriched by specifying the name and associated hand for each finger in contact, and also the order in which they have to enter in contact with the surface.

In figure 2, we represent an illustration of the use of the taxonomy with *Sticky Tools*. Each line represents a mode for the technique, annotated with the number of fingers associated, and with either a "*d*" for direct or "*i*" for indirect. Indirect fingers are represented with an "*i*" in the corresponding circles whereas direct fingers are left blank. Circles connected together with a single-line represent DOF of the task controlled in an integral way. Groups of circles disconnected represent DOF of the task controlled in a separable way.

For the *Sticky Tools* technique represented in Figure 2, the mode $1d$ represents the first finger in contact with the object to manipulate, which controls the object translation along the screen plane in a direct and integral way. When a second finger enters in contact with the same object (mode $2d$), translation and rotation around the z-axis are now possible in a direct and integral way in addition to the DOF controlled by the first finger (i.e. each finger movement can now change four DOF at the same time). The second finger can also be used in an indirect way (mode $1d$+$1i$) to control two DOF in rotation in an integral way but separately from the DOF controlled by the first finger. Last the mode $2d$+$1i$ shows the combination of the previous modes to control the six degrees of freedom at the same time but with two DOF in rotation being controlled separately.

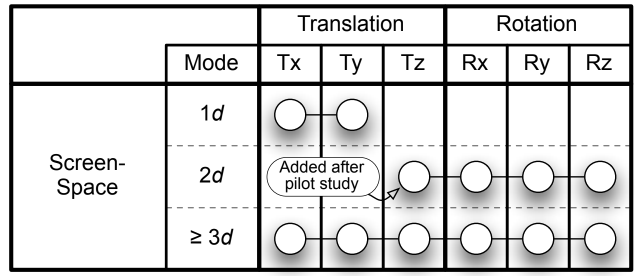

Figure 3: *Description of the Screen-Space technique using the taxonomy.*

4 Screen-Space technique

As mentioned previously, Reisman et al. [2009] introduced a method to perform 3D manipulation with multi-touch displays in an integral way. We refer to this technique as *Screen-Space*. The method consists in solving constraints given by users' fingers. A constraint solver minimizes the error between contact points' screen-space projection and their corresponding screen-space target positions. A simplified version of the algorithm can be described by the following steps :

1. When a finger touches a 3D object projected on screen :
   - Record the 2D location of the finger on screen (point $F2d_1$)
   - Record the 3D point corresponding to the ray-casting of the finger 2D position into the 3D scene (point $P3d_1$)
2. When a finger moves:
   - Record the new position (point $F2d_2$)
   - Use the constrain solver to adjust the position and orientation of the 3D object so that when $F2d_2$ is casted into the scene, it points to $P3d_1$

The goal of the algorithm is to match users' fingers to 3D points and keep these 3D points *stuck under* users' fingers when they move. When it comes to scale and rotate a 2D picture using multi-touch input, it is exactly the same process but instead of matching 2D points (i.e. fingers) with 3D points (i.e. 3D object), 2D points (i.e. fingers) are matched with 2D points (i.e. 2D picture).

To control the six DOF required for 3D manipulation, at least three fingers are required, as a single finger can only control two DOF at best (we consider here only the x,y positions of fingers). With less than three fingers, the interface designer has to choose the DOF controlled in the task. With one finger the natural choice is to control the translation of the object in the camera plane, as illustrated with mode $1d$ in Figure 3. With two fingers, up to four DOF among the six can be controlled. Reisman et al. do not recommend any particular mapping. The mode $2d$ present one possible mapping chosen after a pilot study presented hereafter.

5 Introducing *DS3*

According to Garner [1974], the perceptual structure of 3D manipulation consists of six integrated DOF. Jacob et al. [1994] recommends to match the perceptual structure of the task with the control structure of the input device. Strictly following these two recommendations leads to using only input devices with six integrated

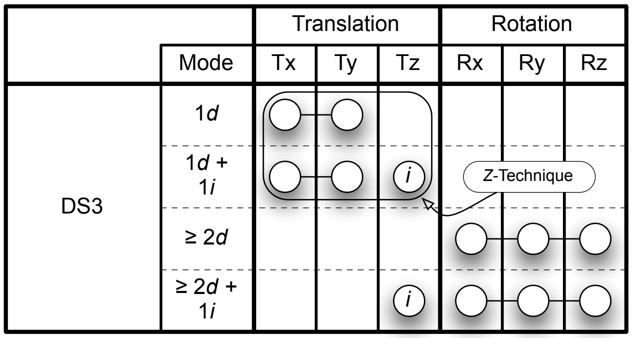

Figure 4: *Description of the DS3 technique using the taxonomy.*

DOF such as 3D mice or 3D wands and to interact with the whole hand instead of only interacting with fingers.

Considering the separable structure of fingers [Ingram et al. 2008], it appears impossible to strictly match the perceptual structure of the task with the control structure of multi-touch displays. The previous work above shows there is no clear answer to this problem. One hand *Screen-Space* proposes to control the six degrees of freedom in an integral way and on the other hand *Sticky Tools* proposes a separation between the degrees of freedom. As these two techniques were not evaluated or compared, it is difficult to know which approach is the best. If the DOF separation appears better, it also addresses the question of the best way to separate DOF.

During an informal evaluation of *Sticky Tools*, we observed that the integral control of translation and rotation (modes $2d$ and $2d{+}1i$ in Figure 2) appears difficult to handle. When DOF are controlled separately, our hypothesis is that a clear separation of translation and rotation improves user efficiency. As a consequence we designed a new technique clearly separating the control of rotation from the control of position. We called this technique *DS3*. It combines the Z-technique [Martinet et al. 2010] used to perform 3D positioning only and the control of rotations with the use of the constraint solver described by [Reisman et al. 2009].

With one direct finger, objects can be translated along the screen plane (mode $1d$ in Figure 4). Depth translation is performed in an indirect way, with a second finger. When this finger is in contact with the surface we measure its relative motion on the surface and use backward forward movement to control the depth position. Forward movement moves the object away from the user view and backward movement moves it closer to the user's view. With at least two direct fingers, users can control the orientation of the object in an integral way using the constrain solver previously described.

The number of fingers directly in contact with the object (one vs. more than one) provides a clear separation between translation and rotation. In addition, when rotating the object, we also allow the manipulation of the object depth (i.e. translation along z-axis) with an indirect finger, as previously described. This is not a breach of the separation of position and orientation as depth-position is handled via an separated additional finger.

6 Pilot experiment

This pilot experiment was designed to pre-test *Sticky Tools*, *Screen-Space* and *DS3* on a real task. We also wanted to experiment different mappings of two fingers with *Screen-Space*. In addition, we wanted to tune some parameters of the techniques.

With *Screen-Space*, we can control up to 4 DOF with 2 fingers. To remove unintended translation as mentioned by [Hancock et al. 2007], we decided to remove the two DOF which were mapped to the one finger mode, leaving us with the four remaining DOF. As mentioned earlier, we believe that separating rotation can improve efficiency. Thus, we decided to also map the two finger mode with the control of rotation DOF only.

6.1 Apparatus

The experiment was conducted on an Immersion iLight[1] touch table based on the Diffused Illumination technique and consisting of a 100 cm $\times$ 70 cm (42 inches) monoscopic display positioned at a distance of 105 cm from the floor. The video-projector under the table was set at 60 Hz with a 1400 $\times$ 1050 pixels resolution giving a pixel density of 14 pixels per cm (36 DPI). A camera running at 120 Hz with a 640 $\times$ 480 pixels resolution and positioned under the surface recorded the finger movements on the surface. This gives a maximum resolution of 6.4 dots per cm (16.25 DPI) for finger tracking. We used the iLight framework version 1.6 for fingers detection and tracking. Fingers data were then sent using TUIO messages[2] to a custom built 3D application based on the Ogre3D framework[3].

6.2 Task and participants

The task is a three dimensional manipulation task based on the docking task introduced by Zhai [Zhai and Milgram 1998]. Each experimental trial began after the previous target was successfully positioned and ended with the successful positioning of the current target. Participants were asked to dock a rendered molecule into a transparent matching molecule-shaped dock as quickly as possible. The control of both position and orientation was necessary to complete the trial. The source molecule was made of five atoms of different colors. Each atom independently turned green when positioned correctly inside the corresponding atom within the error margin. The trial was considered as fully completed when all the atoms stayed at the correct position for 0.8 s. The transparent molecule then moved to another position while the source molecule was repositioned at the center of the screen. In addition to perspective and occlusion, we also added a ground with shadows projection to improve depth perception. The camera remained fixed during the whole experiment. We controlled for the presence of depth (translation along z axis required or not), the combination of axis required for the rotation and the amount of rotation required.

3 males with a mean age of 25 participated. Participants had variable experience with virtual reality and multi-touch displays. One was an expert, another had some experience, and the last one was a novice.

6.3 First results and discussion

Task completion time is defined as the time it takes to successfully position the current molecule into the destination from the last successfully positioned molecule. Mean task completion time was subject to high variability. In addition, results also exhibited a strong learning effect. These results indicate that we should run more blocks (i.e. 3 blocks were present in the pilot study) in the final study to take into account the learning effect.

Users reported that the task was very difficult to perform. They all found the task difficult to grasp due to the presentation of the

[1] http://www.immersion.fr
[2] http://tuio.org
[3] http://www.ogre3d.org

molecule. They reported difficulties in distinguishing rotations and translations needed to perform the task. The 3D perception was also affected because of the object symmetry: in our case, looking at the shadows did not help users as it normally should do. All users reported fatigue and difficulties to manipulate objects when they were located at the top of the screen (i.e. far away from them).

Another surprising comment from users concerned *Screen-Space*. They all complained about depth-translation: they were frustrated to not be able to control the depth position with two fingers. As our mapping controlled orientation only (i.e. 3 DOF), one extra DOF was available by the constraint solver (i.e. 2 fingers allow to control up to 4 DOF). We therefore decided to change our two fingers mapping and we added the control of depth-position in addition to rotation (figure 3).

As a consequence, we designed a new experiment. We decided to increase the number of blocks to five. We also changed the mapping of two fingers with *Screen-Space* to control both depth-position and orientation. Finally we switched to a peg-in-hole task (Figure 1) for two reasons: to remove the symmetry of the object and to position the hole in a place that reduced the fatigue highlighted previously by users (i.e. the hole was located closer to users).

7 Controlled Experiment

7.1 Goals

The main goal of the experiment is to evaluate the effect of DOF separation on multi-touch displays for 3D manipulation. A second objective is to compare *Sticky Tools* and *Screen-Space* which have never been compared or evaluated.

In designing the experiment, we formulated the following hypothesis :

(H1) Based on the results of the pilot study and user feedback, we hypothesize that separating the control of translation from rotation will increase performance as users will not get confused controlling both at the same time.

(H2) Separating the control of translation from rotation will increase coordination (in translation or rotation): if users can manipulate DOF for translation or rotation separately, they will be able to improve the coordination of the DOF for the translation or the rotation.

(H3) The presence of depth-translation will affect performance and coordination, especially with techniques that map depth-translation and rotation together, emphasising the problem pointed out by Nacenta et al. [2009].

7.2 Task

The task is a three dimensional peg-in-hole task based on the one exposed by Unger et al. [2002] (Figure 1) without collision detection enabled. Each experimental trial began after the previous peg was successfully positioned and ended with the successful positioning of the current peg. Participants were asked to position and orientate as quickly as possible a peg into a hole located at the middle of a 3D pavement. The peg was made of a rectangular base on which a cylindrical shape was extruded. The peg color turned from brown to yellow when it was well oriented. When both position and orientation were under a certain threshold, the peg turned to green to indicate it was successfully located. The trial was considered as fully completed when the peg stayed at the correct position during 0.8s. The peg then moved to another position, selected randomly on a semi-sphere (i.e. the center of the semi-sphere was the hole and the radius was defined so as to fit the screen properly). The hole remained at the same place. In addition to perspective and occlusion, we also added a ground with shadows projection to improve depth perception. The camera remained fixed during the whole experiment. The hardware setup was the same as in the pilot study.

7.3 Participants

Eight males and one female with a mean age of 24 (SD 1.5) participated. 6 were right-handed and 3 were left-handed and all had normal or corrected to normal vision. Participants had a variable experience with virtual reality and 3D applications but this is acceptable as we are observing a lower level physical behaviour. Three of them were frequent users of multi-touch displays. The others were familiar with tactile technology such as tactile mobile phone or tablet-PC but never worked for a long time on such a device.

7.4 Design

A repeated measures within-subjects design was used. The independent variables were TECHNIQUE, PRESENCE OF DEPTH, ROTATION LEVEL and ROTATION AMOUNT. There were three levels for TECHNIQUE : *DS3*, *Sticky Tools* and *Screen-Space*. The presentation order of TECHNIQUE was counter-balanced across participants. The PRESENCE OF DEPTH variable had two levels indicating the need, or not, to adjust depth-position. The two levels were NODEPTH and DEPTH. There were also two levels for ROTATION LEVEL, influencing the type of rotation to be performed SIMPLE and COMPLEX. SIMPLE sets the rotation only on one axis (*x*, *y* or *z*) which was randomly chosen. COMPLEX sets the rotation to be a random mix of x, y and z axis. There were two levels as well for ROTATION AMOUNT, changing the angle of the rotation to performed: SMALL and LARGE. For SMALL, the total amount of rotation was to set to an angle of $30°$ while it was $120°$ for LARGE.

As suggested by our pilot experiment, we added extra blocks to the experiment. Participants thus completed five successive BLOCKS of trials. Each BLOCK consisted of 16 trials: 2 repetition of 8 PRESENCE OF DEPTH $\times$ ROTATION LEVEL $\times$ ROTATION AMOUNT combinations. The presentation order of TECHNIQUE was counter-balanced across participants. A break was encouraged after each set of 10 trials.

Before starting the experiment with a new technique, participants had a 15 minutes training period to get used to the current technique. The experiment ended with a qualitative feedback from the participants. The experiment lasted approximately 100 minutes in total.

In summary, the experimental design was : 9 participants $\times$ 3 TECHNIQUES $\times$ 5 BLOCKS $\times$ 2 PRESENCE OF DEPTH $\times$ 2 ROTATION LEVEL $\times$ 2 ROTATION AMOUNT $\times$ 2 repetitions = 2,160 total trials

8 Results

8.1 Task completion time

Task completion time is defined as the time it takes to successfully position the current peg into the hole from the last successfully positioned peg.

Repeated measures analyses of variance found a significant main effect for BLOCK ($F_{4,32} = 20.29$, $p < 0.001$) which shows the presence of a learning effect. Pairwise comparisons show significant differences ($p < 0.05$) between the first three blocks and the

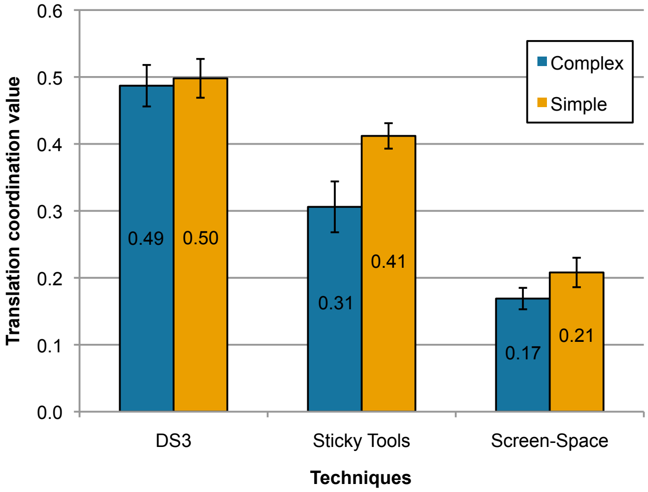

Figure 5: *Mean translation coordination for each technique under the different levels of* ROTATION LEVEL*. Error bars represent 95% confidence interval.*

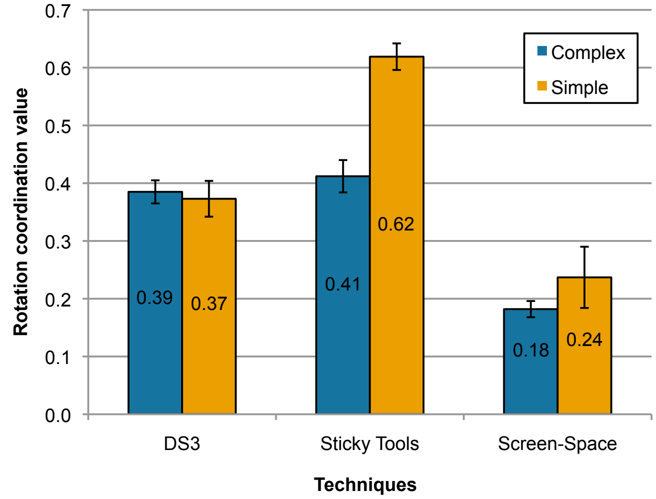

Figure 6: *Mean rotation coordination for each technique under the different levels of* ROTATION LEVEL*. Error bars represent 95% confidence interval.*

last two. As a result we removed the first three blocks for subsequent analysis.

We found a significant main effect of TECHNIQUE on the task completion time ($F_{2,16} = 38.10$, $p < 0.001$). Pairwise comparisons show significant differences ($p \leq 0.030$) between *DS3* (8.85s), *Screen-Space* (18.36s) and *Sticky Tools* (11.16s). This supports **H1** showing that *DS3* is 52% faster than *Screen-Space* and 21% faster than *Sticky Tools*.

As hypothesized (**H3**), we found a significant main effect of PRESENCE OF DEPTH on the task completion time ($F_{1,8} = 8.47$, $p = 0.023$) reducing the mean completion time from 13.44s to 12.14s with no depth adjustment. We also found a significant main effect of ROTATION AMOUNT ($F_{1,8} = 63.32$, $p < 0.001$) reducing the mean completion time from 15.43s to 10.14s with SMALL level.

8.2 Translation Coordination

The translation coordination coefficient ($TrCoord$) is defined as the ratio of the length of the shortest path ($ShortPath$) and the length of the actual path ($RealPath$) [Zhai and Milgram 1998].

Repeated measures analyses of variance found a significant main effect for BLOCK ($F_{4,32} = 4.96$, $p = 0.004$) which shows the presence of a learning effect. Pairwise comparisons show significant differences ($p < 0.05$) between the first block and the others. As a result we removed the first block for subsequent analysis.

We found a significant main effect of TECHNIQUE on translation coordination ($F_{2,16} = 58.22$, $p < 0.001$). Pairwise comparisons show significant differences between *DS3* and *Sticky Tools* ($p = 0.005$), between *DS3* and *Screen-Space* ($p < 0.001$) and finally between *Sticky Tools* and *Screen-Space* ($p < 0.001$). The mean translation coordination is 0.49 for *DS3*, 0.36 for *Sticky Tools* and 0.19 for *Screen-Space*. This result supports **H2**, *DS3* allowing more translation coordination.

As expected (**H3**) we found a significant main effect of PRESENCE OF DEPTH on translation coordination ($F_{1,8} = 132.43$, $p < 0.001$). More interestingly, we found a significant interaction between PRESENCE OF DEPTH and TECHNIQUE ($F_{2,16} = 19.10$, $p < 0.001$). Under the NODEPTH level, *DS3* significantly outperformed *Sticky Tools* ($p = 0.009$) and *Screen-Space* ($p < 0.001$). There was no significant difference between *DS3* and *Sticky Tools* for the other levels where depth-translation was needed ($p = 0.059$).

ROTATION LEVEL also shows a significant main effect on translation coordination ($F_{1,8} = 12.32$, $p = 0.010$). Complex rotations reduce the mean translation coordination from 0.37 to 0.32. A significant interaction effect between ROTATION LEVEL and TECHNIQUE was found ($F_{2,16} = 6.179$, $p = 0.012$). Pairwise comparisons found that *DS3* had no significant change in translation coordination, no matter the type of rotation asked. For *Screen-Space* and *Sticky Tools*, increasing the complexity of the rotation resulted in a significant change in translation coordination, ($p = 0.025$ for *Sticky Tools* and $p = 0.008$ for *Screen-Space*) (Figure 5).

Finally, repeated measures analyses of variance found a significant main effect for ROTATION AMOUNT ($F_{1,8} = 55.71$, $p < 0.001$), larger rotation reducing the mean translation coordination from 0.40 to 0.29.

8.3 Rotation Coordination

The rotation coordination coefficient ($RotCoord$) is defined as the ratio of the initial rotation mismatch ($InitRot$) and the amount of actual rotation ($TotalRot$) [Zhai and Milgram 1998].

We did not find a significant main effect for BLOCKS. Repeated measures analyses of variance found a significant main effect for TECHNIQUE ($F_{2,16} = 43.07$, $p < 0.001$). Pairwise comparisons show significant differences ($p < 0.005$) between *Sticky Tools* and the other techniques. Mean rotation coordination was 0.52 for *Sticky Tools*, 0.38 for *DS3* and 0.21 for *Screen-Space*. This contradicts in part **H2** since *Sticky Tools* outperforms *DS3* for rotation coordination.

As predicted (**H3**), we found a significant main effect of PRESENCE OF DEPTH on rotation coordination ($F_{1,8} = 8.90$, $p = 0.02$). Mean rotation coordination was 0.39 under the NODEPTH level and 0.35 under the DEPTH level.

In addition, we found a significant main effect of ROTATION LEVEL on rotation coordination ($F_{1,8} = 18.58$, $p = 0.004$), complex rotation reducing the mean rotation coordination from 0.41 to 0.32. Interestingly, we also found a significant interaction between ROTATION LEVEL and TECHNIQUE ($F_{2,16} = 23.30$, $p < 0.001$). Pairwise comparisons found that changing the complexity of the rotation significantly changes the mean rotation coordination for *Sticky Tools* ($p = 0.001$) and *Screen-Space* ($p = 0.035$). Surprisingly, the complexity of rotation did not significantly change the mean rotation coordination for *DS3* (Figure 6). In addition, pairwise comparisons did not show any significant difference between *Sticky Tools* and *DS3* with complex rotations.

Finally, we found a significant interaction between PRESENCE OF DEPTH and ROTATION AMOUNT on rotation coordination ($F_{1,8} = 16.66$, $p = 0.005$). Pairwise comparison revealed that when depth-translation is required, increasing the amount of rotation significantly changes the mean rotation coordination. This significant effect was no present under the NODEPTH level.

8.4 Qualitative Feedback

Among the participants, 6 preferred *DS3*, 2 rated *DS3* and *Sticky Tools* equally, and 1 preferred *Sticky Tools*. The participants who said they prefer *DS3* found the technique easier to use and appreciated the fact that rotation was decoupled from translation. According to them, this allowed much more accurate control. They also mentioned that performing rotation was easy and felt *natural*. The participant who prefered *Sticky Tools* found the technique less difficult to use for manipulation with both hands. They also reported that they were able to do everything with only one hand. This is something they strongly prefered even if both hands were required from time to time. In contrast to the other participants, they did not like the sequentiality of *DS3*.

Specifically about *Sticky Tools*, one user reported that the way to handle rotation was efficient but not intuitive. Nevertheless they did not like the fact that depth-translation and rotation were linked together. This difficulty came from the coupling of translation and rotation — this was also pointed out by five other participants. Another user reported that the use of an indirect (i.e. not on the 3D object) finger to control orientation was easier to use, in contrast to *DS3* where the external finger controls the depth-position.

Regarding the *Screen-Space* technique, all participants reported that this technique was very difficult to use to perform the task. two participants reported that, although the technique felt intuitive, it was very difficult to perform specific operations. Four participants liked the technique when working on a plane surface such as the top of the peg. They successfully used a gesture highlighted by Reisman et al. , where four fingers manipulate a 3D plane. Nevertheless, they also pointed out the limitation due to the object size: *Screen-Space* is difficult to use when the size of the object is reduced. Another participant commented that although complex movements were relatively easy to achieve, simple movements were difficult. This was supported by two participants who described the technique as *unpredictable*. Six reported that the integration of all DOF together made the technique difficult to use.

9 Discussion

We designed the experiment to compare three different techniques for performing 3D manipulation tasks on multi-touch displays. *Screen-Space* controlled the six DOF of the task in an integral way, whereas *Sticky Tools* and *DS3* separated the DOF of the task using different strategies.

DOF Separation and Performance

Results show that, for the techniques studied, DOF separation improves performance compared to DOF integration for a docking task on multi-touch displays: both *Sticky Tools* and *DS3* showed significant lower task completion time compared to *Screen-Space*. *DS3* improves performance by 52% compared to *Screen-Space* and *Sticky Tools* by 39%.

This result can be explained by the translation and rotation coordination values which illustrate the relative effectiveness for controling the different DOF during a task. *Screen-Space*, which tightly couples rotation and translation, revealed the lowest translation and rotation coordination. *Sticky Tools* significantly improves translation coordination by 89% and rotation coordination by 148% compared to *Screen-Space*. *DS3* improves translation coordination by 158% and rotation coordination by 81% compared to *Screen-Space*.

It appears that for the integral 3D manipulation task we considered, trying to use the separated DOF of a multi-touch display in an integral way provides lower performance. Instead, separating the DOF of the task to match the separated structure of the input device leads to better results.

This conclusion extends the work of Veit et al. [2009] who found similar results for an orientation task.

DOF Separation Strategies

The experiment showed a significant lower task completion time for *DS3* compared to *Sticky Tools* with a 21% improvement for *DS3*. It shows that the strategy of separating DOF can have a severe impact on performance.

Garner [1974] showed that orientation and position were two integral attributes of 3D manipulation, making the theory of Jacob et al. [1994] difficult to apply to multi-touch displays. However, orientation and position are still two *different* attributes which users can easily separate. The lower completion time for *DS3* supports this idea.

In designing *DS3*, we clearly separated translation from the rotation DOF, leading to a superior completion time and also to higher translation coordination. Providing the control of orientation and depth-positioning at the same time, *Sticky Tools* has a significantly lower translation coordination than *DS3* where no depth-translation is required. We did not find this difference when depth-translation was required, highlighting the fact that users had difficulties performing rotations only when depth-position was present for *Sticky Tools*. This supports our choice to separate the control of translation from the control of rotation.

The interaction between ROTATION LEVEL and TECHNIQUE also support this separation. Combining rotation and translation together, both *Sticky Tools* and *Screen-Space* reduced translation coordination while also increasing the complexity of the rotation task. With a clear separation, the translation coordination of *DS3* was not affected by the complexity of the rotation. This result is also true for rotation coordination. *DS3* was not affected, whereas the two others techniques have reduced coordination: there is no difference between *Sticky Tools* and *DS3* with complex rotations. We believe this result is also related to the interaction between ROTATION AMOUNT and PRESENCE OF DEPTH: when depth-translation is required, the rotation coordination increases with the amount of rotation.

These conclusions extends the work of Nacenta et al. [2009] to 3D manipulation. With techniques mapping depth translation and rotation together, it becomes more difficult to perform large rotations without also affecting position, resulting in poor performance and coordination. Introducing *DS3* with the separation of the control of rotation and translation, we increased performance, coordination, and user satisfaction.

Direct vs. Indirect Control

One key point with *DS3* is the mode switching between the control of rotation and translation. In our case, we differentiate rotation from translation according to the number of finger contacts. However, this may have a major drawback when the size of the object being manipulated is small. For example, using three fingers to perform 3D manipulation with *Screen-Space* is difficult with small objects. Indirect control represents a way to solve this issue.

Indirect control provides users with the ability to clearly separate DOF, even with small objects. However, a drawback is the possibility of inadvertently select or manipulate another nearby object. When designing interaction techniques for multi-touch displays, interface designers should either prevent interaction with multiple objects at once, or provide clear feedback to show which objects users can interact with.

10 Conclusion and Future Work

We have introduced a taxonomy to classify 3D manipulation techniques for multi-touch displays. The study of 3D interaction techniques in relation with the structure of multi-touch displays led us to introduce *DS3*, a 3D manipulation technique based on a total separation of the control of position and orientation. Results of a controlled experiment show the strength of the new technique for both performance and user preference.

Separating translation DOF from rotation DOF had a severe impact and led to the best performance. This relationship between separation of DOF and performance confirms recent results showing that the simultaneous manipulation of all DOF does not necessary lead to the best performance [Veit et al. 2009]. Our study revealed that the integration of both translation and rotation reduce performance, coordination and user satisfaction. A conclusion which extends the work of Nacenta et al. [2009], who covered similar issues for 2D manipulation.

As future work, we plan to further investigate the link between separation of DOF and coordination over time, especially the effect on performance. We also wish to explore the design-space introduced by the taxonomy presented in this paper.

Acknowledgments

This work was supported by the ANR project n°ANR-09-CORD-013 "InSTInCT". We thank the study participants for their time and Daniel Vogel for proof reading.

References

CARD, S. K., MACKINLAY, J. D., AND ROBERTSON, G. G. 1991. A morphological analysis of the design space of input devices. *ACM Trans. Inf. Syst. 9*, 2, 99–122.

GARNER, W. 1974. *The Processing of Information and Structure*.

HANCOCK, M. S., CARPENDALE, S., VERNIER, F. D., WIGDOR, D., AND SHEN, C. 2006. Rotation and translation mechanisms for tabletop interaction. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, 79–88.

HANCOCK, M., CARPENDALE, S., AND COCKBURN, A. 2007. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1147–1156.

HANCOCK, M., CATE, T. T., AND CARPENDALE, S. 2009. Sticky tools: Full 6dof force-based interaction for multi-touch tables. In *Proceedings of Interactive Tabletops and Surfaces 2009*.

HILLIGES, O., IZADI, S., WILSON, A. D., HODGES, S., GARCIA-MENDOZA, A., AND BUTZ, A. 2009. Interactions in the air: adding further depth to interactive tabletops. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, 139–148.

HINCKLEY, K., TULLIO, J., PAUSCH, R., PROFFITT, D., AND KASSELL, N. 1997. Usability analysis of 3d rotation techniques. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, 1–10.

INGRAM, J. N., KÖRDING, K. P., HOWARD, I. S., AND WOLPERT, D. M. 2008. The statistics of natural hand movements. *Experimental Brain Research. 188*, 2, 223–236.

JACOB, R. J. K., SIBERT, L. E., MCFARLANE, D. C., AND MULLEN, J. M. P. 1994. Integrality and separability of input devices. *ACM Trans. Comput.-Hum. Interact. 1*, 1, 3–26.

MARTINET, A., CASIEZ, G., AND GRISONI, L. 2010. The design and evaluation of 3d positioning techniques for multi-touch displays. In *3D User Interfaces, 2010 IEEE Symposium*, 115 –118.

NACENTA, M. A., BAUDISCH, P., BENKO, H., AND WILSON, A. 2009. Separability of spatial manipulations in multi-touch interfaces. In *Proceedings of Graphics Interface 2009*, 175–182.

REISMAN, J. L., DAVIDSON, P. L., AND HAN, J. Y. 2009. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, 69–78.

REKIMOTO, J. 2002. SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, 113–120.

UNGER, B. J., NICOLAIDIS, A., THOMPSON, A., KLATZKY, R. L., HOLLIS, R. L., BERKELMAN, P. J., AND LEDERMAN, S. 2002. Virtual peg-in-hole performance using a 6-dof magnetic levitation haptic device: Comparison with real forces and with visual guidance alone. In *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 263.

VEIT, M., CAPOBIANCO, A., AND BECHMANN, D. 2009. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, 51–58.

WANG, Y., MACKENZIE, C. L., SUMMERS, V. A., AND BOOTH, K. S. 1998. The structure of object transportation and orientation in human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 312–319.

WILSON, A. D., IZADI, S., HILLIGES, O., GARCIA-MENDOZA, A., AND KIRK, D. 2008. Bringing physics to the surface. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, 67–76.

ZHAI, S., AND MILGRAM, P. 1998. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 320–327.