Beyond the mouse: Understanding user gestures for manipulating 3D objects from touchscreen inputs

Aurélie Cohé, Martin Hachet

PII: S0097-8493(12)00156-2
DOI: http://dx.doi.org/10.1016/j.cag.2012.09.004
Reference: CAG2297

To appear in: Computers & Graphics

Received date: 29 June 2012
Revised date: 8 September 2012
Accepted date: 24 September 2012

Cite this article as: Aurélie Cohé and Martin Hachet, Beyond the mouse: Understanding user gestures for manipulating 3D objects from touchscreen inputs, Computers & Graphics, http://dx.doi.org/10.1016/j.cag.2012.09.004

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Beyond the mouse: Understanding user gestures for manipulating 3D objects from touchscreen inputs

Aurélie Cohé, Martin Hachet
INRIA, 351 cours de la Libération, 33405 Talence cedex, France
Université de Bordeaux - CNRS (LaBRI)

Abstract

Multi-touch interfaces have emerged with the widespread use of smartphones. Although a lot of people interact with 2D applications through touchscreens, interaction with 3D applications remains little explored. Most 3D object manipulation techniques have been created by designers who have generally put users aside from the design creation process. We conducted a user study to better understand how non-technical users tend to interact with a 3D object from touchscreen inputs. The experiment consists in 3D cube manipulations along three viewpoints for rotations, scaling and translations (RST). Sixteen users participated and 432 gestures were analyzed. To classify data, we introduce a taxonomy for 3D manipulation gestures with touchscreens. Then, we identify a set of strategies employed by users to perform the proposed cube transformations. Our findings suggest that each participant uses several strategies with a predominant one. Furthermore, we conducted a study to compare touchscreen and mouse interaction for 3D object manipulations. The results suggest that gestures are different according to the device, and touchscreens are preferred for the proposed tasks. Finally, we propose some guidelines to help designers in the creation of more user friendly tools.

Keywords: 3D, touchscreen, human factors, user study, user-centered design

1. Introduction

Touchscreens have been commonly used by the general public since smartphones and tablets have appeared. Therefore, people are used to navigating in 2D maps or photos using touch inputs. On the other hand, 3D graphics applications are still limited due to the difficulty of 3D interaction from 2D inputs. Most of these applications are 3D video games where the user interacts with virtual game pads displayed on screen.

Recently, 3D user interfaces based on touch gestures have been proposed (e.g., [1, 2, 3, 4, 5, 6]). They explore various mapping between finger gestures and corresponding actions in the 3D environment. However, these user interfaces have been designed without any formal investigation on the way non-expert users tend to interact with 3D objects displayed on touchscreens. In this paper, we present a comprehensive user study to better understand the link between how a 3D action is perceived by a user, and the corresponding gesture that is intuitively associated with it.

After reviewing the related work (Section 2), we present our experimental protocol (Section 3.1). In Section 3.2, we introduce a taxonomy based on the analysis of the recorded gestures. We investigate strategies that subjects tend to use for 3D object manipulations on touchscreens (Section 3.3). Then, we focus on common behaviors that led us to define a gesture vocabulary (Section 4). In Section 5, we study how users tend to manipulate 3D objects with a mouse for comparing it with touchscreen results. Finally, we present a general discussion on how users interact with 3D objects, propose a set of design guidelines, and discuss differences between touchscreen and mouse interaction for 3D object transformations (Section 6).

2. Related work

2.1. 3D manipulation on touchscreens

Various techniques have been proposed to manipulate 3D objects with touchscreens from multi-finger inputs. A first set of studies is based on a several finger combination to manipulate multiple degrees of freedom (DOF) simultaneously. Reisman et al. [6] propose a multi-finger co-located technique by introducing a constraint set, formulated as a quadratic problem, to disambiguate user inputs. Hancock et al. introduce a technique that relies on interaction in shallow-depth with one to three fingers [1] or in gravity-based 3D environments [2]. Martinet et al. [5] suggest translating virtual objects in the screen plane by sticking them under one finger, and by using another finger in an indirect way to control their position in depth. In their approach, rotations are performed from two finger inputs using the constraint solver described in [6]. By separating rotations and translations, Martinet et al. [4] conclude that the DOF separability improves easiness and performance. Similarly, Kin et al. [3] develop a gesture set that allows the disambiguation of the intended transformation when manipulating 3D objects. These works introduce new techniques or new gesture sets to manipulate 3D objects. In most of them, a posteriori experiments are conducted to assess the validity of the approach. On the other hand, these techniques have been designed without taking into account how users tend to interact with 3D objects from touch inputs, a
priori. Our goal is to investigate such a priori behaviors for manipulating 3D objects.

2.2. Understanding user gestures
Rather than setting an arbitrary gesture vocabulary, another set of studies is based on the observation of non-guided gestures for the most appropriate gesture corpus definition. Wobbrock et al. [7] build up a study to understand how users interact on a surface with standard control actions such as deleting, copying, pasting, and for picture manipulations. Epps et al. [8] study how users exploit hand shapes to apply standard control actions (i.e., cut, zoom). Koskinen et al. [9] study user preferences and associations on hand movements to understand what is natural and comfortable. Wu et al. [10] develop a set of design principles for gestures applied on touchscreens and then perform a user study allowing to validate or invalidate them. In a 3D context, Cohé et al. [11] create a box-shaped widget for 3D object manipulations on touchscreens. They disambiguate transformations by proposing three different inputs: dual fingers for scaling, one finger movements for rotations, and precise selections for translations. The iTBox widget conception is partially based on an a priori study that has been conducted to determine the natural user gestures to spin a cube. All these works have resulted in some design guidelines based on user behaviors understanding in a specific context. In the same way, we are interested in understanding user preferences specifically for 3D object manipulations on touchscreens.

2.3. Classifications and taxonomies
Some researchers propose to classify the gestures drawn on a surface. Roudaut et al. [12] introduce MicroRolls gestures, defined by the velocity of the tangential force of the skin with the screen surface to distinguish rolls and slides. Wobbrock et al. [7] propose a taxonomy based on hand forms, gesture nature (i.e., symbolic, metaphorical), binding and flow to classify user gestures for action tasks (i.e., open, delete) and 2D transformations (i.e., move, rotate). North et al. [13] classify gestures on surfaces for a 2D object selection task. In 3D, Martinet et al. [4] introduce a taxonomy to classify translation and rotation techniques by representing the relationship between the finger number, their directness and the controlled DOF. Similarly to these works, we aim at analyzing the gestures during a 3D object manipulation task. In particular, we are interested in classifying these gestures.

3. 3D object manipulation study with a touchscreen

3.1. Experimental protocol

3.1.1. Apparatus
The experimental environment is composed of a TouchCo 13 inches-sized multi-touch surface that records input data and of an Optima video-projector with a resolution of 1280 × 800 pixels for the display (see Figure 1). Half of the image is projected on an interaction zone and the second half is projected on a visualization zone. The projector is set perpendicularly to the table to minimize image distortion.

Figure 1: Experimental setup.

3.1.2. Procedure
For the experiment, the following procedure is applied. First, the participant is asked to sit in front of the interaction zone and to adjust the chair. Then, a three-second video shows an object transformation on the visualization zone and the user is asked to draw a gesture on a static image displayed on the interaction zone (the first image of the video) that matches the best, according to her/him, this transformation. We tell the participants that the action they should take for the given transformation is off the top of their head and they can use the same gesture for different transformations. Finally, the participant is asked to assess her/his gestures with two statements, similarly to [7]:

- “the gesture I did is a good match for its intended purpose” (QT1),
- “the gesture I did is easy to perform” (QT2).

Two seven-level Likert scales are used to evaluating these statements with a ranking from strongly disagree (1) to strongly agree (7). The process is repeated for a set of 27 pre-recorded videos, in a random order. At the end of the experiment, we interview each user to obtain additional feedback.

The 27 videos illustrate 3D transformations (rotation, translation, scaling) on a basic shape (a cube) along the three object frame axes, and for three viewpoints, as shown in Figure 2. Rotations are performed counterclockwise with an angle of 90 degrees. For scaling, the applied factor is 150%. Translation directions are the same as those of the axes in Figure 2 and their displacements are twice the size of the cube. The projected size of the smallest visible edge is 3.9 centimeters. In the 3D scene, lighting is enabled with a directional light located at the top of the scene and the object shadow is projected on the ground for a better perception of the object displacement. For front view, the shadow helps disambiguating translation from scaling in depth for instance. Note that we call translation in depth and scaling in depth the cases when the transformation is applied along an axis aligned with the viewing direction. Static images are used in the interaction zone to prevent users from picking an anchor point and following its trajectory during the transformation. Consequently, the study is focused on the way users control inputs on touchscreens without being influenced by outputs.
3.1.3. Participants

We set up the system in a scientific museum and asked visitors if they were willing to participate in the experiment. Sixteen subjects volunteered, four men and twelve women aged from 19 to 60 (mean = 31.56, standard deviation (σ) = 10.31). All of them were right-handed and all have little or no experience manipulating virtual 3D objects or touchscreens.

We chose novice users to limit the influence of the gestures acquired from specific learned tools. Participants ranked their touchscreen habit with a mean score of 1.92 (σ = 1.44) and 1.37 (σ = 0.68) for their 3D habit (using five-level Likert scales from never use (1) to everyday use (5)). Participants came from various professional backgrounds, including teaching, saling, managing, electrical engineering and so on. We collected 27 gestures per user, corresponding to the tested video sequences, that is 432 gestures in total. Eleven of them were considered as erroneous due to inappropriate recordings linked to the touch surface sensitivity.

3.2. Classifying gestures

3.2.1. Definition

Wobbrock et al. [7] presented a taxonomy for surface gestures applied in a 2D context or for standard control actions. They classified them in four parameters: form, nature, binding, and flow. Form distinguishes static or dynamic pose and path for each hand. Nature indicates if the gesture is symbolic, physical, metaphorical or abstract. Binding shows if it is object-centric, world-dependent, world-independent or mixed dependencies. Flow specifies if the response occurs after or while the user is interacting. We were inspired by this taxonomy to study the gestures in a 3D context. Compared to Wobbrock et al. [7], we do not use the same taxonomy because, in our case, all the gestures are physical, object-centric and continuous. Moreover, we do not take into account the form-of-the-hand parameter because we want our results to be valid on any sensing technologies, including those that are not capable of tracking such a parameter.

Consequently, we classify the gestures according to three parameters: form, initial point locations (IPLs) and finger trajectory. Form indicates how many fingers are used and if these related inputs are static or dynamic. A non static finger is defined as a path finger. IPLs describe the locations of the initial inputs on the cube (e.g., corner, edge, face or external to the cube). They also define the relationship between the picked elements and the applied transformation (e.g., the picked edge is orthogonal to the Transformation Axis (TA)).

We consider a vertex as an IPL when the distance between the finger and the vertex is less than half of the average fingertip, which is nine millimeters large according to Dandekar et al. [14]. The same distance is used to determine if an IPL is an edge or a face. Moreover, to describe the position of a face relative to an axis (e.g., the TA), we use the supporting plane of the face as a referent. So, when an IPL is on a face, we consider two cases: the face is orthogonal or parallel to the TA. The trajectory defines if the finger trajectory is along the Transformation Direction (TD) and if there is an intersection between the trajectory and an edge of the cube. Indeed, such intersection points may have an impact on user gestures to spin a cube, as shown in tBox [11]. Figure 3 shows a concrete example illustrating these definitions.

Table 4 summarizes the results for all the transformations, as for each transformation type separately. Note that some cases that have not been used in this study are not illustrated (e.g., when two orthogonal edges of different faces are picked). The results reveal that the vast majority of gestures are applied with one or two path fingers and are applied in the transformation direction. Moreover, it can be observed that subjects tend to initially pick one face. This classification provides a global picture of the gestures performed by the participants. The following sections focus on the correlation between the Form, IPLs and the Trajectory, for each transformation type.

For rotations, the user gesture analysis leads to the emergence of ten different gesture categories, illustrated in Figure 4 (R1-R10):

- gestures for which the trajectory is along the TD (70.92%). These gestures are curved:
  - R1: the IPL is on a face parallel to the TA (17.95%)
  - R2: the IPL is on an edge parallel to the TA (11.48%)
  - R3: the IPLs are on two corners of an edge (10.61%)
  - R4: the IPL is external to the cube and the trajectory intersects the cube (10.05%)
## Illustrations of gesture categories summarized for one viewpoint and for one axis, and this for each transformation type: around the X axis for rotations (left, R1-R10) and along the Z axis for scaling (center, S1-S6) and translations (right, T1-T5).

### Rotation

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td><img src="rotation0.png" alt="Image" /></td>
</tr>
<tr>
<td>0.75</td>
<td><img src="rotation0.75.png" alt="Image" /></td>
</tr>
<tr>
<td>1.5</td>
<td><img src="rotation1.5.png" alt="Image" /></td>
</tr>
<tr>
<td>2.25</td>
<td><img src="rotation2.25.png" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="rotation3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

- R5: the IPL is on a corner (7.96%)  
- R6: the IPL is on a face orthogonal to the TA (6.43%)  
- R7: the IPLs are on two corners of a diagonal of a face orthogonal to the TA (5.00%)  
- R8: the IPLs are on two different faces, one being orthogonal to the TA and the other one being parallel to it (1.43%)  
- R9: gestures that are tangent to a circle corresponding to the rotation centered on a point of the TA at the intersection point between the trajectory and an edge of the cube, which is orthogonal to the TA (total 9.93%, corresponding to 2.13% that are picked on a face parallel to the TA, 1.42% on an edge parallel to the TA and 6.38% that are picked outside the cube)  
- R10: gestures that are tangent to a circle corresponding to the rotation centered on a point of the TA at the picked point (total 6.38%, corresponding to 3.55% that are picked on a face parallel to the TA, 0.71% on a corner and 2.13% on an edge parallel to the TA)

### Scaling

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td><img src="scaling0.png" alt="Image" /></td>
</tr>
<tr>
<td>1.5</td>
<td><img src="scaling1.5.png" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="scaling3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

- S1: the IPLs are on a face parallel to the TA (36.57%)  
- S2: the IPLs are on four corners of a same face parallel to the TA (12.69%)  
- S3: the IPLs are on a same edge parallel to the TA (11.94%)  
- S4: the IPLs are on two faces, one is orthogonal to the TA and the other is parallel to it (10.45%)  
- S5: the IPLs are on three different faces (2.24%)  
- S6: the IPLs are on three different faces (10.45%)  

### Translation

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td><img src="translation0.png" alt="Image" /></td>
</tr>
<tr>
<td>1.5</td>
<td><img src="translation1.5.png" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="translation3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

- T1: the IPLs are on two different faces, one being orthogonal to the TA and the other one being parallel to it (11.94%)

All gestures can be classified in the categories described above. Note that for the straight gesture case, 12.77% of the gestures can be both R9 and R10 as the tangent to the picked point and the tangent to the intersection point are superimposed (7.80% are picked on a face parallel to the TA, 0.71% on a face orthogonal to the TA and 4.26% on an edge parallel to the TA). 11.27% of the gestures are composed of one static finger and one path finger. For the three specific cases where the TA is parallel to the screen, 19 gestures out of 46 have been performed with a straight gesture that has finished with a curved trajectory. This can be explained by the fact that users who performed these gestures wanted to distinguish the rotation from the translation, when the screen projection of the triangles corresponding to the rotation are straight. All the other gestures are straight in these cases.

For **scaling operations**, as illustrated in Figure 4, six major categories appear:

- S1: the IPLs are on a face parallel to the TA (36.57%)  
- S2: the IPLs are on four corners of a same face parallel to the TA (12.69%)  
- S3: the IPLs are on a same edge parallel to the TA (11.94%)  
- S4: the IPLs are on two faces, one is orthogonal to the TA and the other is parallel to it (10.45%)  
- S5: the IPLs are on two opposite edges orthogonal to the TA and on a same face parallel to the TA (8.21%)  
- S6: the IPLs are on three different faces (2.24%)
Note that two-path-finger gestures correspond to two-joint-finger gestures, and the two paths are identified similarly by their IPL and their trajectory (see Section 4.2). Moreover, all gestures are along the TD, except those which are performed for the scaling in depth. Considering the form, 3.70% of the scaling gestures have been performed with only one finger. This corresponds to a side effect of the perspective, i.e., when only one face seems to be translated during the movement (Y axis for view (b) in Figure 2). Thus, we consider only IPLs to detect categories.

Except for scaling in depth gestures, only four gestures belong to none of the categories (2.99%): one had IPLs on two faces parallel to the TA, one on three corners, one on a face orthogonal to the TA and another one outside of the cube without intersecting the cube. For the particular case of scaling in depth, all subjects but one picked the only face that is orthogonal to the TA (11.19%). All of them performed a pinch gesture and reported they assimilated this to the zoom functionality for 2D pictures. Interestingly, 78.16% of the gestures have been performed on a face, or corners and edges that belong to a face parallel to the TA.

For translations, five categories emerge (see Figure 4):

- **T1**: the IPL is on a face parallel to the TA and the trajectory is along the TD (34.04%)
- **T2**: the IPL is on a face orthogonal to the TA and the trajectory is along the TD (28.38%)
- **T3**: the IPL is external to the cube and the trajectory is along the TD and intersects the cube (15.65%)
- **T4**: the IPL is on an edge parallel to the TA and the trajectory is along the TD (5.67%)
- **T5**: the IPL is external to the cube and the trajectory is along the TD and does not intersect the cube (4.96%)

Similarly to scaling, two-path-finger gestures are identified as two-joint-fingers (see Section 4.2). Gestures performed with one static finger for translation in depth correspond to gestures along the TD. 11.31% of the gestures cannot be classified: one gesture has been performed with four fingers, two using corners and thirteen other gestures fall into the particular case of translation in depth.

For this ambiguous transformation, a lot of different gestures have been performed for the particular case of translation in depth. One user created her/his own gesture by performing a dual finger gesture from the external to the center of the cube (0.71%), two of them used one static finger gesture on the face (1.42% included in T2), four of them used a pinch gesture (2.84%) and others performed their gesture from bottom to top (5.63%). Considering gestures applied for the translation in depth on front viewed cube, 4.92% of the whole gestures picked the face, 0.71% an edge, 1.42% a corner, 3.55% picked a point external to the cube, and one of them picked the shadow of the cube.

### 3.3. Strategies

#### 3.3.1. Overview

We define strategies as a gesture interpretation of the user mental models. This interpretation comes from observations, as well as from participant interviews. In a preliminary study, Cohé et al. [11] detected two different strategies to spin a cube with one finger:

- **Grab**: the user picks a point on a cube face and follows a path having in mind the projection of the picked 3D point will remain under her/his finger.
- **Push**: the user follows a trajectory tangent to the intended rotation at the point where this trajectory collides with the projection of an edge, as if she/he were pushing the cube from this edge.

One goal of this paper is to generalize previous results with an in-depth analysis of user strategies for rotations, as for scaling and translations. Interestingly, no user tried to use an existing interaction method, such as the virtual trackball [15], probably because all the participants were novice for 3D application.

#### 3.3.2. Categories of gestures

In this section, we compare the strategies based on their relevancy (QT1) and easiness (QT2) scores. For each comparison, we use Chi square tests ($\chi^2$) with two variables: strategies (e.g., A part, Extremities, Dual grab and Others for scaling operations) and scores (low or high). For scores, we consider as low scores those lower than the median score of the evaluated parameter (relevancy or easiness) and as high scores those higher than it. However, when the participant number in each cell of the contingency table is composed of less than five gestures, the Fisher exact probability test ($pF$) is employed. Moreover, a Bonferroni correction test is performed when more than two strategies are concerned, to compare them two by two.

In our experiment, user strategies described for rotations in [11] can be identified. R1 is similar to the Grab strategy and R9 to the Push strategy. For both of them, additional properties linked to the trajectory path characteristics can be identified:

- **Curved**: the gesture trajectory is curved. The user follows the trajectory of the picked point or the intersection point between the finger trajectory and the cube.
- **Straight**: the trajectory is straight. The user throws the cube. Therefore, the trajectory is tangent to the picked point for the Grab strategy and it is tangent to the intersection point between the finger trajectory and a cube edge for the Push strategy.

Consequently, we redefine Grab and Push as follow:

- **Grab**: the user picks a point on the cube surface and then moves the object.
- **Push**: the user begins her/his gesture and the cube moves after it has been pushed (i.e., when the finger trajectory intersects an edge orthogonal to the TA).
Furthermore, for all gestures performed with one static finger and one path finger, users, as reported in their interviews, define the transformation axis with the static finger and perform a motion with the second finger. We call this the **Axis** strategy.

Each gesture category can be classified with strategies described above (see Figure 5 for strategy illustrations). R4 can be associated to CurvedAndPush, R9 to StraightAndPush, R10 to StraightAndGrab, and the others to Curved. R1, R2, R4, R5, R6, R9 and R10 can also be assimilated to Axis for gestures performed with one static finger and one path finger.

As reported by users, all gestures for which users picked a point of the cube, and for which the trajectory intersects an edge of the cube, are either **Grab** or **Push** strategies. In these cases, it is not possible to identify clearly which strategy is involved, as the 3D projection of the initial point may be on the cube (Grab) or not (Push). This ambiguity linked to the static aspect of the images occurs on 54.23% of the gestures.

**Figure 5**: Rotation strategies: CurvedAndGrab (a), CurvedAndPush (b), StraightAndGrab (c) and StraightAndPush (d). Blue dot lines indicate when the finger moves while the cube remains static. Red lines indicate when both the finger and the object move. Stars indicate impact points between the finger and the object. Dark points show that the user picks the cube.

Table 1 shows the quantitative results for each strategy. The most used strategy is CurvedAndGrab. We suppose that this is due to the mouse interaction habits, where similar paradigms are used. All other techniques are unusual with a mouse, such as Axis, which requires two actions (one finger defines the axis and another one performs the action), whereas multi-touch enables users to do both at the same time. The **Push** strategy use can be explained by the nature of the gesture, which relies on real life actions. The high scores for relevancy indicate that users have performed these gestures with confidence. Moreover, users are more confident with the **Push** strategy than with the **Grab** one ($\chi^2=5.15, p<0.05$). The scores for easiness reveal that **Straight** is easier to perform than the **Curved** strategy ($\chi^2=3.9, p<0.05$), straight gestures being simpler than curved ones. Considering relevancy, we did not find any significant correlation in using or not the **Axis** strategy.

For most scaling operations, three different strategies can be defined:

- **A part**: the user scales a part of the cube on a scaled face.
- **Extremities**: the user scales the cube using opposite edges of a scaled face.
- **Dual grab**: the user picks two points of several cube elements and performs a dual finger gesture in the scaling direction.

Table 1: Rates, mean of relevancy and easiness for each rotation strategy. For each strategy implying **Grab** or **Push**, the percentages are expressed on gestures where there are no ambiguities (i.e., 45.77% of rotation gestures).

Table 2 shows the statistical results for each strategy. The vast majority of gestures relies on the **A part** strategy. We make the assumption that this technique favors an easy selection compared to the **Extremities** strategy, which requires precise selection. This hypothesis is reinforced by the results obtained for question QT2 on easiness. Unlike the **Dual Grab**, **A part** and **Extremities** strategies are performed on a unique face of the cube. The high scores for relevancy and easiness indicate that users have done these gestures with confidence and without any difficulty.

Table 2: Rates, mean of relevancy and easiness for the different scaling strategies.

Three strategies appear from the observations made on translations (see Figure 6 for strategy illustrations):

- **Push**: the user begins her/his gesture outside the cube and draws it towards the object as if she/he was pushing it.
- **Without object referent**: the user performs a gesture outside the cube and the trajectory is straight and along the TD, without intersecting the cube.
- **Grab**: the user picks a point of the cube and follows the translation trajectory. It includes three sub-strategies:
  - **Lateral**: the user picks a point of the lateral face and follows its trajectory.
− **Pull:** the user picks a point of the pulled face and follows its trajectory.

− **Push:** the user picks a point of the pushed face and pushes the cube.

![Translation strategies](image)

Figure 6: Translation strategies. Blue dot lines indicate when the finger moves while the cube remains static. Red lines indicate when both the finger and the object move. The star indicates the impact point between the finger and the object. The dark point shows that the user picks the cube.

T3 is included in the **Push** strategy, T1 and T4 in the **GrabLateral** one, T5 in the **Without objec referent** one and T2 in the **GrabAndPush** one or in the **GrabAndPull** one according to the picked face. One gesture is outside the categories defined for translations. This gesture was performed to the bottom for translating the face to the the top. The subject who performed this gesture said he wanted to bounce the cube. Similarly to the main strategies, this behavior can be linked to its real life counterpart.

Table 3 shows statistical results about these strategies. The user relevancy depends on the strategy ($\chi^2 = 18.50, p < 0.005$). The **Without Object Referent** strategy is perceived as less relevant than the **GrabLateral** one ($pF < 0.05$) and the **GrabAndPush** one ($pF < 0.01$). The **Others** category, corresponding to unclassified singular gestures, is less relevant than the **GrabAndPush** one ($pF < 0.01$) and the **Push** one ($\chi^2 = 8.23, p < 0.005$). Similarly to rotations, the **Grab** strategies are more used than other strategies and, as suggested before, we assume that this observation related to the user habits with the mouse. Interestingly, the relevancy and easiness scores for the **Push** strategy are higher than for the **Grab** strategies. We suppose that this could be due to the fact that the **Grab** strategies require more precise selections.

<table>
<thead>
<tr>
<th>Strategies</th>
<th>%</th>
<th>Relevancy mean</th>
<th>Relevancy σ</th>
<th>Easiness mean</th>
<th>Easiness σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Push</td>
<td>15.65</td>
<td>6.24</td>
<td>1.04</td>
<td>6.43</td>
<td>0.68</td>
</tr>
<tr>
<td>W/o object R.*</td>
<td>4.96</td>
<td>4.86</td>
<td>1.07</td>
<td>5.29</td>
<td>0.49</td>
</tr>
<tr>
<td>GrabLateral</td>
<td>40.42</td>
<td>5.92</td>
<td>0.96</td>
<td>6.17</td>
<td>0.82</td>
</tr>
<tr>
<td>GrabAndPull</td>
<td>3.55</td>
<td>5.80</td>
<td>1.30</td>
<td>6.00</td>
<td>1.00</td>
</tr>
<tr>
<td>GrabAndPush</td>
<td>24.82</td>
<td>6.11</td>
<td>0.76</td>
<td>6.17</td>
<td>0.86</td>
</tr>
<tr>
<td>Others</td>
<td>10.56</td>
<td>5.07</td>
<td>1.83</td>
<td>5.67</td>
<td>1.40</td>
</tr>
</tbody>
</table>

Table 3: Rates, mean of relevancy and easiness for the different translation strategies. (* Without object referent)

### 3.4. Additional strategy analysis

Several strategies are used for each transformation type (rotations, scaling or translations). Figure 7 shows the distribution of the most applied participant strategy. The analysis of this distribution shows that most of the participants tend to use a predominant strategy for all their gestures in the same transformation type, even if several strategies for each transformation type are used. We also found that each strategy are applied by at least half of the participants, except the **Axis** strategy for rotations, the **Extremities** and the **Dual grab** ones for scaling and the **Without object referent** one for translations.

![Transformation type](image)

Figure 7: Use of the predominant participant strategy. 100% (blue bars) means that a participant uses always the same strategy, ≥ 75% (red bars) means that she/he uses a strategy with at least 75% of her/his gestures, and so on.

Furthermore, no viewpoint effect on the applied strategy is found for rotations with **Grab** and **Push** strategies ($\chi^2 = 0.40, p > 0.8$), with the **Axis** strategy ($\chi^2 = 0.21, p > 0.9$), for scaling ($\chi^2 = 0.08, p > 0.9$) and for translations ($\chi^2 = 0.43, p > 0.8$).

For the **Curved** and **Straight** strategies, post-hoc tests do not reveal effects neither.

### 4. Common behaviors

#### 4.1. Single parameter analysis

In the previous section, we analyzed each gesture as a whole. In the following, we focus on three gesture parameters that we analyze independently.

Trajectories for each transformation type A key point for an interaction technique is to detect *a priori* transformations that users intend to apply. Scaling are easy to differentiate from translations and rotations as these transformations are applied with two or four fingers in opposite directions. The main observed difference between translations and rotations is the gesture trajectory. Most of the rotation gestures are curved (72.54%), whereas translation gestures are straight (98.92%). Nevertheless, some rotation gestures are straight (27.46%) and some gestures seem to be straight at the beginning and circular at the end because the projection of the circle defining the rotation is elliptic. Therefore, real time detection of what the user wants to do at the beginning of the gesture may be hard to infer. However, some strategies differ for translations and rotations (*i.e.*, **Without object referent** and **Axis**). For **Axis**, we can detect *a priori* that the user intends to apply a rotation from the positions of the two fingers. For **Without object referent**, we cannot detect the transformation type *a priori* because this action can, in fact, be the beginning of a **Push** strategy.

Nevertheless, it is possible to detect *a posteriori* if the gesture trajectory does not intersect the cube.
**Gesture location** 92.5% of the rotation gestures follow a virtual circle corresponding to the related rotation. In the same way, 84.18% of the gestures for scaling are centered close to the related transformation axis and 91% for translations. Therefore, from this observation, we can detect whether a gesture is a translation or a rotation by observing if the gesture is around a translation axis and not around a rotation circle, and vice-versa.

However, it does not clarify the cases where the rotation circle and the translation axis are superimposed (e.g., it happens for the rotation around the X axis and the translation along the Y axis with the viewpoint (b) in Figure 2).

**Number of fingers** Participants do not always use the minimal number of fingers required for a given strategy (e.g., each finger has generally one independent role unlike joint finger gestures). 21.21% of the gestures are performed with at least two fingers that have the same role. Note that this is not only true for one specific transformation type (rates of gestures composed by fingers with similar roles: 17.04% of the gestures for rotations, 23.50% for scaling and 26.98% for translations) nor for a small number of users (9/16 participants have performed this kind of gestures).

### 4.2. User-defined trajectory set

The definition of a gesture set that would fit a large number of users requires the identification of a common behavior that is largely shared. We analyzed if the gesture trajectory can be used to this purpose. If several fingers have the same trajectory, only one of these fingers is taken into account (see Figure 8).

![Figure 8: Definition of trajectories taken into account. For (b) and (c), the gestures contain only one trajectory which moves the object and which corresponds to the trajectory applied on the gesture (a). So, we consider the trajectory (b) and (c) similar to the (a) trajectory.](image)

**Common behavior** The trajectory agreement for each video is evaluated using the formula introduced by Wobbrock et al. [16]. We simplify this formula as we consider the agreement for a given transformation and not for a transformation set. The agreement score is given by:

$$\sum_{P_t \subseteq P} \left( \frac{|P_t|}{|P_t|} \right)^2$$

In our analysis, $r$ is a transformation, $P_r$ is the proposed trajectory set for the transformation $r$, and $P_t$ is a subset of identical trajectories from $P_r$. The range for the agreement is $[|P_r|^{-1}, 1]$. For instance, for the rotation around the X axis from the viewpoint (a) in Figure 2, six users drew the same trajectory, and ten drew another one. The agreement for the corresponding video is thus $(6/16)^2 + (10/16)^2 = 0.53$. For the whole set of gestures, the agreement about trajectories is high: 20 trajectories have an agreement greater than 0.7 (mean = 0.95, $\sigma = 0.09$), six trajectories between 0.5 and 0.7 (mean = 0.56, $\sigma = 0.07$) and one trajectory agreement is equal to 0.4 and corresponds to translation in depth for front viewed cube. These results indicate that an interaction technique based on trajectories could be relevant. One major advantage of using such an interaction technique is that it takes into account most of the gestures that users draw naturally, without the need to learn an arbitrary language.

**Ambiguities** We consider that two gesture trajectories are in conflict if they are similar for two different transformations with the same viewpoint. For instance, for the viewpoint (b) in Figure 2, four gesture trajectories are in conflict considering the trajectory from the bottom to the top, as it is used for rotation around the X axis, for scaling along the Y axis and for translations along the Y and Z axes. In our experiment, 16 trajectories are in conflict: nine for the viewpoint (a) in Figure 2 and six for the viewpoint (b). Note that the more visible the faces, the less conflicts in trajectories ($\chi^2 = 58.5, p < 0.0001$). The maximal coverage of trajectory set without any ambiguity is equal to 83.57% of the whole gestures. Table 6 shows the user-defined trajectory set, the user agreements and the ambiguities.

### 5. Comparison with mouse gestures

#### 5.1. Experimental protocol

We conducted a second user study to compare previous results with gestures applied with the mouse. The goal is to know how users tend to naturally interact with a mouse for 3D object manipulations and to answer the following questions:

- Are gestures the same with the mouse and the touchscreen for the 3D object manipulation? (QM1)
- Do users prefer touchscreens or mice for the proposed transformations? (QM2)

The same apparatus have been used to avoid parameter influence due to any modifications between the two setups. The touch surface has just been replaced by a mouse with two buttons and one scroll wheel button. The same videos as for the first experiment have been used and the same protocol has been applied. We explained the users that they were able to use the scroll wheel in the both directions and click on one button or several ones at a time. We also said them that if they wanted to create a path, they had to maintain a button pressed all along the movement.

The experiment took place in the same museum. Ten subjects (three women and seven men) aged from 24 to 64 ($mean = 44.10, \sigma = 11.29$) participated. All of them had little or no experience manipulating virtual 3D objects ($mean = 2.1, \sigma = 1.45$ with a five-level Likert scale from never use (1) to everyday use (5)) and all used the mouse everyday. As in the first experiment, users came from various professional backgrounds. The experiment lasted about 15 minutes per participant.
5.2. Classifying gestures

5.2.1. Taxonomy dimension analysis

Taxonomy definition First, we analyze mouse gestures to characterize them according to the taxonomy we defined for touch gestures (see Section 3.2.1). So, in the same manner as before, we consider an element as the IPL when its distance from the finger is less than half of the average fingertip. In addition to this taxonomy, we add a dimension describing the buttons activated by the user. Table 5 shows the results according to this taxonomy. We observe that the wide majority of gestures are applied pressing the left button only for all transformation types and using the scroll wheel for scaling operations. Furthermore, we observe that the trajectory is a path along the TD for translations and rotations, and to the TD or one of the opposite directions of the TDs for scaling operations.

Note that the form can contain several paths as the gesture can be performed with several steps: the user can maintain a button when she/he moves the mouse, releases it and performs this action again as it is considered as a single gesture.

Impact of the transformation type To understand user mouse behaviors, we analyzed the correlation between each taxonomy dimension and the transformation types.

For the button dimension, the transformation types have no significant effect on the button choice. The left button is used for almost all gestures and interviews reveal that subjects never gave importance to the button they used. However, transformation types significantly affect the use of the scroll wheel ($\chi^2 = 21.74, p < 0.0001$), which is more used for scaling operations than for translations ($pF < 0.001$) and rotations ($pF < 0.05$). Indeed, participants who used the scroll wheel for scaling operations said they associated the action with the zoom functionality.

Moreover, transformation types affect the form ($\chi^2 = 73.29, p < 0.0001$) and both one point and one path gestures are more used for translations ($pF < 0.001$) and rotations ($pF < 0.001$) than for scaling. Indeed, some participants used several paths for scaling operations to represent the two directions of scaling operations. Nevertheless, some of them used only one path to represent only one direction of the two opposite directions, but they reported a frustration feeling.

No meaningful link was found between transformation types and IPL. However, transformation type affects not only the gesture direction (i.e., along the TD or not ($\chi^2 = 148, p < 0.0001$)), but also if there is an intersection or not between the object and the trajectory ($\chi^2 = 11.60, p < 0.01$).

These results indicate that buttons, forms and trajectories mainly differ with scaling operations from other transformation types. Nevertheless, translations and rotations need a deeper analysis to distinguish them from each other.

Comparisons with touch gestures The results we obtained indicate that the device affects the number of path or points ($\chi^2 = 85.85, p < 0.0001$), the picked element and the TA ($\chi^2 = 20.98, p < 0.0001$). No effect was found on the gesture trajectory intersection with the object.

For form, multi-path gestures are less applied with the mouse than with the touchscreen. These gestures seem to be more natural to do with touchscreens because they can be applied in one step only, which is not possible with the mouse.

For IPL, we observe that faces are more used than gestures with IPL on edges with the touchscreen compared to the mouse ($pF < 0.01$). Gestures with IPL external to the cube or without IPL are more used than faces ($pF < 0.001$) and edges ($pF < 0.001$) with the touchscreen than the mouse. Results are similar with corners, which are more used than edges with the touchscreen than with the mouse ($pF < 0.001$). The fact there are less gestures with IPL external to the cube with the mouse indicates that users prefer to pick a point on the cube, which is close to the Grab strategy, commonly used with the mouse. Note that no significant difference was found between the number of picked elements parallel and orthogonal to the TA. The only difference is that elements neither parallel nor orthogonal to the TA are more used than parallel ones ($pF < 0.05$) and orthogonal ones ($pF < 0.05$) with the touchscreen than with the mouse. This result is linked to the fact there are few gestures with IPL external to the cube with the mouse, contrary to the touchscreen.

Finally, for Trajectory, many gesture trajectories correspond to only half of the TD for scaling operations with the mouse, contrary to the touchscreen where they applied two or four path fingers in the TD.

These first comparisons indicate a change in gestures according to the device for each dimension independently. However, the main differences seem to be due to the nature of scaling operations that are symmetric, while mice provide only one 2D input, contrary to touchscreens.

5.2.2. Classification

Here, we analyze the gesture as a whole. Figure 9 shows our gesture classification realized in the same way as the one described in Section 3.2.2 for touch gestures.

For rotations, all gestures except one can be classified with previous classes (R1:18.89%, R2:15.56%, R5:13.33%, R6:14.44%, R9orR10:25.55%) plus one additional class that defines gestures with IPL on an edge orthogonal to the TA and for which the trajectory is along the TD (R11, 11.11%).

For scaling operations, 65.17% of gestures fall into four main classes: one whose gestures have IPLs on a face parallel to the TA and with a trajectory in both directions (S1, 20.22%), one whose gestures have an IPL on an edge orthogonal to the TA and with a trajectory in one direction of the scaling operation (S5.2, 20.22%), one whose gestures have IPLs on a face parallel to the TA and with a trajectory in one direction of the scaling operation (S1.2, 14.61%) and one whose gestures have two static points on opposite edges of a face parallel to the TA combined with the scroll wheel inputs (S7, 10.11%). Minor classes correspond to gestures with IPL on an edge parallel to the TA and with a trajectory in one direction of the scaling operation (S3.2, 5.62%) and gestures corresponding to S4 class (S4, 4.49%). All other classes have less than three gestures.
Thus, there is a distinct difference for gestures according to the nature of the device for scaling operations. It was predictable because scaling operations in the experiment are symmetric.

For translations, all gestures except seven can be classified in T1 (46.59%), T2 (23.86%), T4 (4.55%) and a new class defined by one path along the TD and the IPL is on an edge orthogonal to the TA (T6, 17.05%).

5.2.3. Strategies

Here, we identify which strategies are used per transformation with the mouse. For rotations, all gestures are performed in picking one point of the cube. Moreover, 74.45% of the gestures follow the Curved strategy, and straight gestures are both tangent to the picked point and to the intersection point as these tangents are superimposed. The straight gestures are performed only when the projection of the rotation circle centered on the cube is straight. User interviews reveal that subjects did not want to push the cube, but they always wanted to grab a point and follow its direction. It is totally different compared to strategies applied with the touchscreen, where 41.54% of gestures correspond to the Push strategy. For translations, all gestures also correspond to the Grab strategy, contrary to the previous results with the touchscreen.

For scaling operations, we observe three new strategies:

- Scroll wheel: the user indicates the transformation axis using fingers and scale using the scroll wheel (21.34% of gestures, including those of S7).
- Asymmetric gesture: the user performs the scaling operation picking one point and drawing its trajectory all along the transformation. The user supposes the system performs the same operation symmetrically (46.01% of gestures, including those of S1.2, S3.2 and S5.2).
- Several steps: the user indicates both the opposite direction of the scaling operation using several moves (30.33% of gestures, including those of S1 and S4).

As the touchscreen strategies cannot be performed with the mouse, comparing with the mouse strategies is not consistent. The Several steps strategy could be assimilated to the ones of the touchscreen but the difference is that users perform several non-symmetrical scales.

We compare the whole rates of relevancy and easiness for the touchscreen and the mouse (see Figure 10). Participants using the touchscreen perceived their gestures as more relevant (χ² = 9.34, p < 0.01 for rotations, χ² = 26.74, p < 0.0001 for scaling, χ² = 6.43, p < 0.05 for translations) and easier (χ² = 16.08, p < 0.0001 for rotations, χ² = 20.13, p < 0.0001 for scaling, χ² = 6.48, p < 0.05 for translations) than those using the mouse. The difference is more important with scaling operations (the Phi coefficient of association is 0.34 for relevancy (φQT1) and 0.30 for easiness (φQT2)) than with rotations (φQT1 = 0.20, φQT2 = 0.26). Moreover, the difference is higher with rotations than with translations (phi QT1 = 0.17, phi QT2 = 0.17). This order is certainly due to the symmetric scaling transformations where users prefer to apply two opposite gestures. For rotations, the fact that users try to apply curved gestures may influence the results, as curved gestures are easier to apply from finger gestures than from mouse inputs.

![Figure 9: Illustrations of mouse gesture categories summarized for one viewpoint and for one axis, and this for each transformation type: around the X axis for rotations (left) and along the Z axis for scaling (center) and translations (right). Rows show gesture categories percentages below how often the corresponding category is used, over all viewpoints and all axes. Green triangles indicate when the IPL is an edge, orange circles when it is a face, blue squares when it is a corner and purple diamonds when it is external to the cube. For S7, the mouse picture indicates participants uses the scroll wheel.](image)

![Figure 10: Scores about relevancy (a) and easiness (b) for touchscreen and mouse gestures.](image)

6. Discussion and conclusion

6.1. User understanding

Many questions about 3D interaction with touchscreens remain unanswered. In this paper, we give first answers for the following questions:
Q1: Is there a common behavior to manipulate a 3D object?

Q2: Are there particular elements of the object with which users interact?

Q3: Do users rely their gestures on strategies such as physically plausible movements, or on a gesture language they create?

Q4: Do users always follow one strategy, or do they use several ones?

Q5: Do they always use the same number of fingers for a given transformation type and strategy?

Q6: Can we deduce a unique gesture set? Is this gesture set conflict-free?

We observed that many strategies are used, but almost all gestures are continuous and are based on physically plausible behaviors, and on object-centric movements (Q1). Moreover, users do not seem to rely on particular elements for a given transformation type: it depends on the applied strategy (e.g., if a user pushes the cube to make it spin, the gesture intersects the 2D projection of an edge of the object) (Q2). We also discovered that users follow several strategies (Q3) and that each user mainly uses one of these strategies (Q4). We also observed that users do not consider as important the number of fingers they use (Q5), thus confirming Wobbrock et al. observations [7]. Furthermore, users tend to follow a common behavior in a given situation (Q1) and a conflict-free trajectory set defines a wide majority of gestures (Q6). Finally, some strategies are similar to those created by users where fingers are colocated with manipulated points (e.g., scaling operations, strategies and Grab strategies for rotations and translations)

6.2. Comparison between mouse and touch gestures

By comparing the mouse and the touchscreen, we observed that strategies are different both for rotations and translations as all gestures for the mouse rely on a Grab strategy, confirmed by users interviews (QM1). We observed differences in strategies for scaling operations. Indeed, the wide majority of gestures have been done with at least two fingers with the touchscreen, whereas the mouse provides only one 2D input. So, with the mouse, users applied gestures by performing several scaling operations or they applied only one trajectory supposing that the other part of the cube is transformed symmetrically. Other participants associated scroll wheel with scaling operations. In this last case, they said they applied this because they were thinking about the zoom functionality. Finally, an extension of this study with one side scaling could be very interesting.

Moreover, touchscreen users were more confident in their gestures than those with the mouse (QM2). It is probably due to the fact that users felt freer to apply strategies they want, and not only the Grab one. This result might explain that touchscreens are more adapted for 3D object manipulations than mice. It could also be explained by the fact participants of the touchscreen experiment were novices in the use of this device, while those of the mouse experiment were experts in the use of the mouse. Therefore, the fact that users apply the Grab strategy almost all the time can be due to an habit, whereas strategies apply with the touchscreen were more “natural” as they were not use to this device.

6.3. Implications for design

We studied the gestures performed in 3D on touchscreens by novice users to understand how they tend to interact. One reason to perform this kind of study is to find invariant behaviors in user gestures. Hence, it may help in the design of new 3D user interfaces, where interaction techniques can benefit from a good understanding of the user perception for the possible actions. From these results, we propose the following guidelines:

- **Favor physically plausible interaction.** A wide majority of gestures of this study rely on physically plausible gestures.

- **For a wide use, favor the Grab strategy for rotations and translations, and the A part strategy for scaling** if you choose only one strategy per transformation type and you want to maximize the number of gestures that are taken into account. The related strategies have been the most used for each transformation type.

- **For easiness, favor the Straight strategy for rotations, the Push strategy for translations and the A part strategy for scaling** if you choose only one strategy per transformation type.

- **For a vast use and to support several strategies, favor interaction techniques based on gesture trajectory analysis.** As described in Section 4.2. 83.57% of the gestures of this study can be identified (transformation type and axis) by their trajectory. Some ambiguities can be clarified with a join analysis of the initial point locations.

According to the mouse study results, we propose these additional guidelines:

- **Adapt interaction techniques according to the device.** The results indicate that users tend to interact differently depending on the input device to perform the proposed transformations. However, we should perform additional studies with other devices to confirm this conclusion as user expertise with the mouse could affect these differences.

- **Favor Grab strategy for 3D object manipulations with mouse.** Participants mainly use a Grab strategy for translations and rotations. Interestingly, users are not used to interacting in 3D contexts with the mouse, but they are used to applying the Grab strategy in 2D contexts.
Moreover, the mouse study indicates an effect of transformation types on button, form and trajectory dimensions. It indicates that these dimensions could help to distinguish which transformation users intend to apply.

6.4. Limitations and future work

This study has been conducted on static images to focus on the way users interact with touchscreen inputs. However, it could be interesting to take into account outputs to verify whether these gestures are still valid when the object moves. Moreover, rotations and translations have been performed according to one direction per axis only, for minimizing the duration of the study and, consequently, the user fatigue. Nevertheless, the use of the dominant or non-dominant hand may impact the results, thus an extended study should be performed to analyze this impact. Furthermore, our hypothesis was that the viewpoint influences the strategy choice. We did not manage to validate this hypothesis in this study. A new analysis dedicated to the viewpoint influence would definitely extend our results.

In our work, basic transformation types on a basic shape and with local axes have been studied to understand user gestures. In a first step. It would be interesting to extend this study with other axes to know if there is a general behavior for all axes. An extension with objects more complex than a cube is necessary to explore the influence of parameters such as curvature. This would allow the generalization of the results to a wider spectrum of 3D objects. Indeed, the strategies defined in this paper rely on elements defining the transformation axis, like R3 where the user picks two points on an edge parallel to the transformation axis. However, this selection is impossible with more complex objects, as the Stanford bunny. In these case, a box-shaped widget, as the one of tBox [11], can be used. In the same manner, extending this study to physical properties and to the object size (e.g., to know if more fingers are used when the virtual object looks heavier or bigger) would be valuable. Moreover, an extension with different screen sizes would be interesting. Indeed, the screen size have strong influence on the way people interact. Indeed, in our experiment, the size enables participants to use several fingers and hands without arm movements, contrary to small phone touchscreen or larger touch surface. The influence of the object position on the strategy choice could also be explored. For instance, if the object is located at a boundary of the screen, users may prefer the GrabAndPull strategy to translate the object to the middle of the screen. The same exploration could be done for scenes with several objects. It could also be interesting to extend this study by allowing subjects to try out two or three different gestures to determine if a user always uses the same strategy in the same context. Moreover, some ambiguities have been detected in the trajectory analysis. One solution to disambiguate these situations could be to find another parameter that most of the users would control in the same way. For instance, the quantity of force applied on the touch sensor could be interesting to investigate. Finally, a human performance study could be interesting to confirm or infirm that user generated gestures outperform pre-defined gestures.

This study has been conducted to better understand user gestures for a 3D manipulation task. We hope these results will help to design new interaction techniques dedicated to touchscreens.

Acknowledgements

We wish to thank the reviewers, Jérôme Baril, Romain Panowski and Pauline Davignon for their help. We also thank Cap Sciences, the museum where the experiment took place. This work was supported by the ANR project ANR-09-CORD-013 InSTInCT - http://anr-instinct.cap-sciences.net.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Values for each parameter</th>
<th>total (%)</th>
<th>R (%)</th>
<th>S (%)</th>
<th>T (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Buttons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Right button</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Without scroll wheel</td>
<td>1.87</td>
<td>0.00</td>
<td>2.25</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>With scroll wheel</td>
<td>0.37</td>
<td>1.12</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Left button</td>
<td>88.76</td>
<td>98.89</td>
<td>76.40</td>
<td>90.91</td>
</tr>
<tr>
<td></td>
<td>Without scroll wheel</td>
<td>6.37</td>
<td>0.00</td>
<td>15.73</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>With scroll wheel</td>
<td>0.37</td>
<td>0.00</td>
<td>1.12</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Wheel button</td>
<td>2.25</td>
<td>1.11</td>
<td>3.37</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>Scroll only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 static point</td>
<td>4.49</td>
<td>1.11</td>
<td>8.99</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>4 static points</td>
<td>0.38</td>
<td>0.00</td>
<td>0.00</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>1 path finger</td>
<td>79.03</td>
<td>97.78</td>
<td>47.19</td>
<td>92.05</td>
</tr>
<tr>
<td></td>
<td>2 path fingers</td>
<td>12.36</td>
<td>0.00</td>
<td>37.08</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>4 path fingers</td>
<td>1.50</td>
<td>0.00</td>
<td>3.37</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>none</td>
<td>2.25</td>
<td>1.11</td>
<td>3.37</td>
<td>2.27</td>
</tr>
<tr>
<td><strong>Form</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 face</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>orthogonal to the TA</td>
<td>16.10</td>
<td>14.44</td>
<td>8.99</td>
<td>25.00</td>
</tr>
<tr>
<td></td>
<td>parallel to the TA</td>
<td>40.07</td>
<td>34.44</td>
<td>40.45</td>
<td>45.45</td>
</tr>
<tr>
<td></td>
<td>both</td>
<td>1.50</td>
<td>0.00</td>
<td>4.49</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>2 faces</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>orthogonal to the TA</td>
<td>17.60</td>
<td>16.67</td>
<td>21.35</td>
<td>14.77</td>
</tr>
<tr>
<td></td>
<td>parallel to the TA</td>
<td>11.24</td>
<td>18.89</td>
<td>7.87</td>
<td>6.82</td>
</tr>
<tr>
<td></td>
<td>1 edge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>orthogonal to the TA</td>
<td>7.11</td>
<td>14.44</td>
<td>3.37</td>
<td>3.41</td>
</tr>
<tr>
<td></td>
<td>parallel to the TA</td>
<td>2.25</td>
<td>1.11</td>
<td>3.37</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Initial Point Locations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 corner</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>same edge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>orthogonal to the TA</td>
<td>55.43</td>
<td>52.22</td>
<td>25.84</td>
<td>88.64</td>
</tr>
<tr>
<td></td>
<td>parallel to the TA</td>
<td>7.49</td>
<td>20.00</td>
<td>1.12</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>both</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 edge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>orthogonal to the TA</td>
<td>13.86</td>
<td>99.00</td>
<td>41.57</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>parallel to the TA</td>
<td>1.87</td>
<td>0.00</td>
<td>5.62</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>none</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trajectory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>along the TD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>intersects the cube</td>
<td>9.73</td>
<td>22.22</td>
<td>4.49</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>does not intersect the cube</td>
<td>1.50</td>
<td>3.33</td>
<td>0.00</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>none</td>
<td>10.11</td>
<td>2.22</td>
<td>21.35</td>
<td>6.82</td>
</tr>
</tbody>
</table>

Table A: Touch 3D gesture classification, rates for the whole transformations and rates for each transformation type (R is rotation, S scaling and T translation).

Table B: Mouse 3D gesture classification, rates for the whole transformations and rates for each transformation type (R is rotation, S scaling and T translation). *for scaling operations only.
Table 6: User-defined trajectory set. Columns represent the different views. Rows correspond to the actions (R is rotations, S scaling and T translations) according to the axis (X, Y and Z). On each of these major rows, the first minor row shows the different trajectories applied for the given transformation type according to the given viewpoint. Under each trajectory, the ratio indicates the number of users who performed this trajectory (“/16”). In the third minor row, the mean of matching (m) (QT1), with standard deviation in parentheses, and the mean of easiness (e) (QT2), with standard deviation in parentheses, and the agreement (a) are given. In each column, the ambiguous trajectories are highlighted by the use of the same color for the same trajectory use on different actions.
<table>
<thead>
<tr>
<th></th>
<th>using a touchscreen</th>
<th>using a mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rotation</strong></td>
<td><img src="image1" alt="Rotation gestures" /></td>
<td><img src="image2" alt="Rotation gestures" /></td>
</tr>
<tr>
<td><strong>Translation</strong></td>
<td><img src="image3" alt="Translation gestures" /></td>
<td><img src="image4" alt="Translation gestures" /></td>
</tr>
<tr>
<td><strong>Scaling</strong></td>
<td><img src="image5" alt="Scaling gestures" /></td>
<td><img src="image6" alt="Scaling gestures" /></td>
</tr>
</tbody>
</table>

**taxonomy**

**user mental models**

**device comparisons**

**design guidelines**
Highlights

- We introduce a taxonomy for 3D object manipulation on touchscreens
- We identify a set of strategies followed by novice users to manipulate a cube
- We highlight how touchscreen interaction differ from mouse interaction
- We show that novices preferred touchscreens than mice for 3D object manipulations
- We propose some guidelines to help the design of new user friendly interfaces