Understanding user gestures for manipulating 3D objects from touchscreen inputs

Aurélie Cohé, Martin Hachet

To cite this version:

Aurélie Cohé, Martin Hachet. Understanding user gestures for manipulating 3D objects from touchscreen inputs. Graphics Interface, May 2012, Toronto, Canada. ACM, pp.157-164, 2012. <hal-00680561>

HAL Id: hal-00680561

https://hal.inria.fr/hal-00680561

Submitted on 26 Jul 2012

Understanding user gestures for manipulating 3D objects from touchscreen inputs

Aurélie Cohé[*] Martin Hachet[†]

INRIA Bordeaux
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 of 3D object manipulation techniques have been created by designers and users are generally put aside from the design creation process. We conducted a user study to better understand how non-technical users interact with a 3D object from touchscreen inputs. The experiment has been conducted while users manipulated a 3D cube with three points of view 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 realize the proposed cube transformations. Our findings suggest that each participant uses several strategies with a predominant one. Finally, we propose some guidelines to help designers in the creation of more user friendly tools.

Index Terms: H.5.2 [Information Systems]: Information interfaces and presentation—Evaluation methodology, interaction styles, user-centered design

1 Introduction

Touchscreens have been commonly used by the general public since smartphones and tablets appeared. Therefore, people are used to navigate in 2D maps or photos from touch inputs. On the other hand, applications that rely on 3D graphics are still limited due to the difficulty of interacting in 3D environments from 2D inputs. Most of these applications are 3D videogames where the user interacts thanks to virtual game pads displayed on screen.

Recently, 3D user interfaces based on touch gestures have been proposed (*e.g.*, [4, 5, 6, 8, 9, 11]). 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 that has been conducted 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). In Section 4, we introduce a taxonomy based on the analysis of the users' gestures. We investigate the strategies that subjects tend to use for the manipulation of 3D objects on touchscreens (Section 5). Then we focus on common behaviors that have led us to define a gesture vocabulary (Section 6). Finally, we present a general discussion about the way users interact with a 3D object and we propose a set of design guidelines (Section 7).

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 works is based on a combination of several fingers to manipulate multiple degrees of freedom (DOF) simultaneously. Reisman *et al.* [11] proposed a multi-finger co-located technique by introducing a set of constraints, formulated as a quadratic problem, to disambiguate user inputs. Hancock *et al.* introduced a technique that relies on interaction in shallow-depth with one to three fingers [4] or in gravity-based 3D environments [5]. Martinet *et al.* [9] suggested to translate virtual objects in the screen plane by sticking them under one finger, and to use 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 [11]. By separating rotations and translations, Martinet *et al.* [8] conclude that the separability of the DOF improves easiness and performance. Similarly, Kin *et al.* [6] developed a set of gestures that allows to disambiguate the intended transformation when manipulating 3D objects. All these works introduced new techniques or new sets of gestures to manipulate 3D objects. Most of them conducted *a-posteriori* experiments 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*. The goal of our work 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 works is based on the observation of non-guided gestures for the definition of the most appropriate gesture corpus. Wobbrock *et al.* [14] built up a study to understand how users interact on a surface with standard control actions such as deleting, copying, pasting, and for the manipulation of pictures. Epps *et al.* [3] studied how users exploit hand shapes to apply standard control actions (*i.e.*, cut, zoom). Koskinen *et al.* [7] studied user preferences and associations on hand movements to understand what is *natural* and *comfortable*. Wu *et al.* [15] developed a set of design principles for gestures applied on touchscreens and then performed a user study allowing to validate or invalidate them. In a 3D context, Cohé *et al.* [1] have created 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 conception of the widget 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 resulted in some design guidelines by understanding user behaviors in a specific context. In the same way, we are interested in understanding user preferences for the specific context of 3D object manipulations on touchscreens.

2.3 Classifications and taxonomies

Some researchers have proposed to classify the gestures drawn on a surface. Roudaut *et al.* [12] introduced MicroRolls gestures, de-

[*] e-mail:aurelie.cohe@inria.fr

[†] e-mail:martin.hachet@inria.fr

fined by the velocity of the tangential force of the skin with the screen surface to distinguish rolls and slides. Wobbrock *et al.* [14] have proposed 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.* [10] have classified gestures on surfaces for a selection task of 2D objects with the number of hands used and the number of items in a group. In 3D, Martinet *et al.* [8] introduced a taxonomy to classify interaction techniques for the translation and rotation of an object by representing the relationship between the number of fingers, their directness (*e.g.*, the distance of each of them with the projection on the screen of the manipulated object) and the controlled DOF. Similarly to these works, in our approach, we aim at analyzing the gestures used during a 3D object manipulation task. In particular, we are interested in classifying these gestures according to the strategies applied by users.

3 Experimental protocol

3.1 Apparatus

The experimental environment is composed of a TouchCo 13 inches-sized multi-touch surface used to record 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.

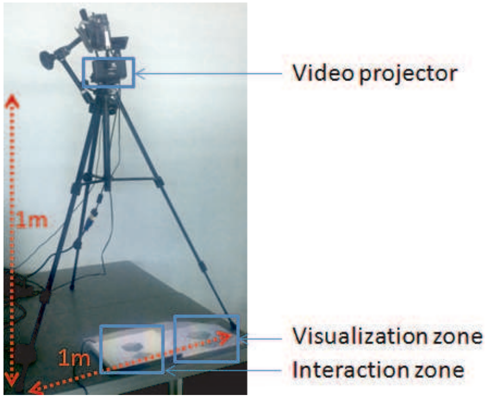


Figure 1: Experimental setup.

3.2 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 him, this transformation. Finally, the participant is asked to assess his gestures with two statements, similarly to [14]:

- "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 were used to evaluate 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 points of view, 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. In the 3D scene, lighting is enabled with a directional light located at the top of the scene and the shadow of the object is projected on the ground for a better perception of the object displacement. For front view, the shadow helps to disambiguate 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, unlike videos of the transformation, to prevent users from picking an anchor point and following its trajectory during the transformation. It allows us to focus the study on the way users control inputs on touchscreens without being influenced by outputs.

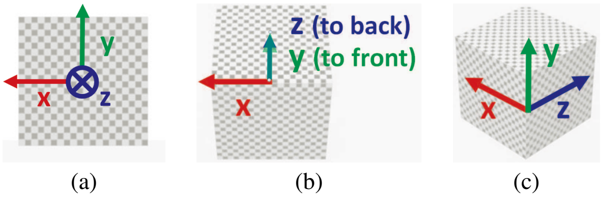


Figure 2: The three viewpoints used during the experiment and the object-centric frames.

3.3 Participants

We set up the system in a scientific museum and asked visitors if they were willing to participate to the experiment. Sixteen subjects volunteered, four men and twelve women aged from 19 to 60 (mean = 31.56, standard deviation (SD) = 10.31). All of them were right-handed and all have little or never manipulated both virtual 3D objects and touchscreens. We chose novice users to limit the influence of gestures acquired from specific learned tools. 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 sensitivity of the touch surface.

4 Classifying gestures

4.1 Definition

Wobbrock *et al.* [14] 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 applied in a 3D spatial context. Compared to Wobbrock *et al.* [14], 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* because we want our results to be valid on any sensing technologies, including those that are not capable of tracking such a parameter.

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.*,

Dimensions	Values for each parameter				total (%)	R (%)	S (%)	T (%)
Form	1 static finger				0.48	0	0	1.42
	1 static finger and 1 path finger				4.03	11.27	0	0.71
	1 path finger				49.81	71.69	3.70	72.73
	2 path fingers				35.76	17.04	66.67	24.43
	4 path fingers				9.93	0	29.63	0.71
Initial Point Locations	faces	1 face	orthogonal to the TA		17.55	7.14	11.94	33.30
			parallel to the TA		34.21	31.43	36.57	34.75
		2 faces	orthogonal to the TA		0	0	0	0
			parallel to the TA		0.24	0	0.75	0
			both		3.85	1.43	10.45	0
		3 faces			0.72	0	2.24	0
	edges	1 edge	orthogonal to the TA		0	0	0	0
			parallel to the TA		12.55	19.29	11.94	6.38
		2 edges orthogonal to the TA*	same face	orthogonal to the TA	0	0	0	0
				parallel to the TA	2.64	0	8.21	0
	corners	1 corner			3.42	8.67	0	1.42
		2 corners	same edge	orthogonal to the TA	3.11	9.18	0	0
				parallel to the TA	0.48	1.43	0	0
			same face (diagonal)	orthogonal to the TA	1.69	5.00	0	0
				parallel to the TA	0	0	0	0
			diagonal of the cube		0	0	0	0
		3 corners			0.24	0	0.75	0
		4 corners	same face	parallel to the TA	4.09	0	12.69	0
	external				14.24	16.43	1.50	24.16
Trajectory	collinear to the TD		intersects the cube		63.14	59.57	46.86	75.81
			does not intersect the cube		22.88	11.35	41.11	12.82
	not collinear to the TD		intersects the cube		13.69	29.09	11.95	10.60
			does not intersect the cube		0.48	0	0.75	0.71

Table 1: 3D gesture classification, rates for the whole transformations and rates for each transformation type (R is rotation, S scaling and T translation). (* To minimize table size, "2 edges" and "orthogonal to the TA" are in a same cell; they should be in two separate columns).

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.* [2]. The same distance is used to determine if an IPL is an edge or a face. Moreover, for comparisons with a face and the TA, we use the supporting plane of the face as a referent. 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 collinear to 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, as it has been shown in tBox [1] to spin a cube. Figure 3 shows a concrete example illustrating these definitions.

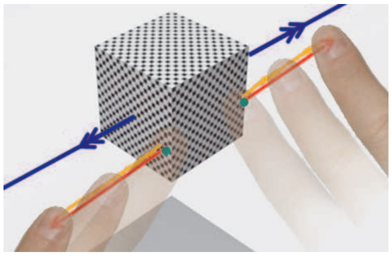


Figure 3: Example of a scaling operation. The blue line is the *Transformation Axis (TA)* and the blue arrows show the *Transformation Direction (TD)*. The green dots are the Initial Point Locations (IPLs) and the red arrows represent the trajectory. The *form* of the gesture is *two path fingers*, the *IPLs* are located on two edges that are orthogonal to the TA and on the same face, which is parallel to this TA. In this example, the finger trajectory is collinear to the TD.

4.2 Classification

Table 1 summarizes the results for all the transformations, as well as for each transformation type separately. Note that some cases that have not be 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 direction of the transformation. 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.

4.2.1 Rotations

The analysis of user gestures for rotation leads to the emergence of ten different categories of gestures, illustrated in Figure 4 (R1-R10):

- gestures for which the trajectory is collinear to the TDs (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%)
  - 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%)

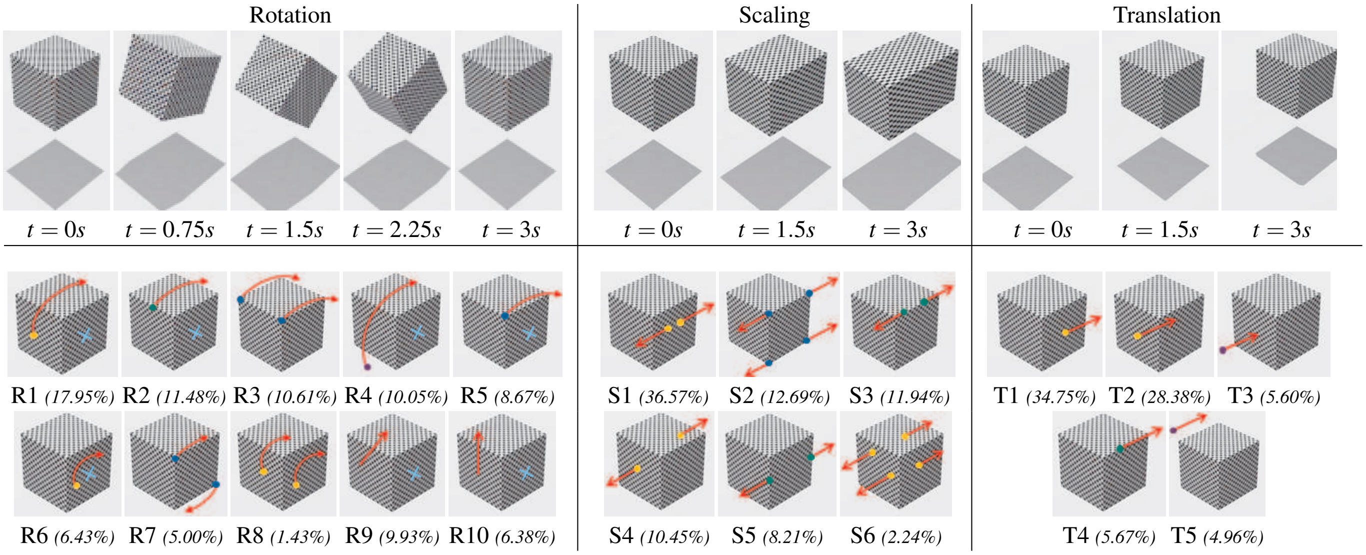


Figure 4: Illustrations of gesture categories summarized for one point of view and for one axis, and this for each transformation type: around the X axis for rotations (left, R1-R10) and along the Z axis for scaling (center, S1-S6) and translations (right, T1-T5). The first row illustrates image samples and the related timeline (t) from the video sequences that were presented to the subjects. The second and third rows show gesture categories and their rates. On these images, the blue crosses are optional static fingers that participants may use. Colored dots indicate the positions of the IPLs. Green is used when an IPL is an edge, orange when it is a face, blue when it is a corner and purple when it is external to the cube. For R9 and R10, no IPLs are specified.

  - R8: the IPLs are on two different faces, one being orthogonal to the TA and the other one being parallel to it (1.43%)

- gestures for which the trajectory is not collinear to the TD (29.09%). These gestures are straight. Both sub-categories are distinguishable in observing more particularly the trajectory:
  - 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)

All gestures can be classified in the categories described above. Note that for the case of straight gestures, 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 particular 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 wanted to distinguish the rotation from the translation, when the projection on the screen of circles corresponding to the rotation are straight. All the other gestures are straight in these cases.

4.2.2 Scaling

For scaling, six major categories, illustrated in Figure 4, 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 in the same manner by their IPL and their trajectory (see Section 6.2). Moreover, all gestures are collinear to 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, corresponding to the side effect of 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 (2.99%) belong to none of the categories: 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.

4.2.3 Translations

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

- T1: the IPL is on a face parallel to the TA and the trajectory is collinear to the TD (34.04%)

- T2: the IPL is on a face orthogonal to the TA and the trajectory is collinear to the TD (28.38%)
- T3: the IPL is external to the cube and the trajectory is collinear to the TD and intersects the cube (15.60%)
- T4: the IPL is on an edge parallel to the TA and the trajectory is collinear to the TD (5.67%)
- T5: the IPL is external to the cube and the trajectory is collinear to 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 6.2). Note that gestures performed with one static finger or four fingers have been applied for *translation in depth* for front view, where static finger gestures correspond to gestures collinear to the TD, and four finger gestures are not classified in the previous categories as they are used for one context only.

Apart from particular gestures used for *translation in depth* with front view, only one gesture does not belong to these categories. This gesture was drawn to the bottom for a translation to the top with the front view of the cube. Moreover, for the particular case of *translation in depth* for front viewed cube, lots of different gestures have been performed. One user created his own gesture in 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 T1), 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, of which one picked the shadow.

5 Strategies

5.1 Overview

We define strategies as an interpretation of the mental models on which users rely their gestures. This interpretation comes from observations, as well as from participant interviews. In a preliminary study, Cohé *et al.* [1] have detected two different strategies to spin a cube with one finger:

- Grab: the user picks a point of a cube face and follows a path having in mind that the projection of the picked 3D point will remain under his finger.
- Push: the user follows a trajectory that is tangent to the intended rotation at the point where this trajectory collides with the projection of an edge, as if 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 well as for scaling and translations. Note that no user try to use an existing interaction method, such as the *virtual trackball*, probably because all the participants were novice users with 3D applications.

5.2 Categories of gestures

In this section, we compare the strategies based on the scores of relevancy (QT1) and easiness (QT2). For each comparison, we divided the participants in two sub-groups: those with a score lower than the median score of the evaluated parameter and those with a score higher than it. We used Chi square tests (χ^2) when sub-groups were composed of at least five gestures. Otherwise, the Fisher exact probability test (pF) was used. Moreover, a Bonferroni correction test was performed when more than two strategies were concerned, to compare them two by two.

5.2.1 Rotations

In our experiment, user strategies described by Cohé *et al.* [1] 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 characteristics of trajectory path 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 and launches the object. 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 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 indicate the transformation axis with the static finger and perform a movement with the second finger, as they reported in the interviews. We call this the ***Axis*** strategy.

Each gesture category can be classified with strategies described above. 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. 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, as reported by user interview. In these cases, it is not possible to clearly identify 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.

Table 2 shows the quantitative results for each strategy. *CurvedAndGrab* is the most used strategy. We suppose that this is due to the habits of mouse interaction, 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 performes the action), whereas multi-touch enables users to do both at the same time. The use of the *Push* strategy 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 (χ^2=5.15, p<0.05). The scores for easiness reveal that *Straight* is easier to perform than the *Curved* strategy (χ^2=3.9, p<0.05), straight gestures being simpler than curved ones. Note that we did not find any significant correlation in using or not the *Axis* strategy for relevancy.

5.2.2 Scaling

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 elements of the cube and performs a dual finger gesture in the scaling direction.

Strategies	%	Relevancy (QT1)		Easiness (QT2)	
		mean	SD	mean	SD
CurvedAndGrab	43.08	5.36	1.16	5.72	1.32
StraightAndPush	21.54	5.29	1.82	6.36	1.08
CurvedAndPush	20.00	6.15	0.90	5.85	1.21
StraightAndGrab	15.38	6.1	0.99	6.6	0.70
Curved	70.92	5.56	1.17	5.76	1.17
Straight	29.09	5.64	1.53	6.33	0.90
Grab	58.46	5.55	1.16	5.84	1.26
Push	41.54	5.70	1.49	6.11	1.15
Without Axis	88.73	5.56	1.28	5.87	1.10
With Axis	11.27	5.81	1.28	6.19	1.33

Table 2: Rates, mean of relevancy and easiness for each rotation strategy.

For *scaling in depth*, the majority of users applied 2D metaphors such as the one used with standard smartphones applications. S1 and S3 are included in the *A part* strategy, S2 and S5 in the *Extremities* one and, S4 and S6 in the *Dual grab* one.

Table 3 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 of the 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.

Strategies	%	Relevancy (QT1)		Easiness (QT2)	
		mean	SD	mean	SD
A part	48.51	6.13	0.82	6.16	0.99
Extremities	20.90	6.15	0.86	5.78	1.50
Dual grab	12.69	6.00	0.79	6.18	0.73
Others	17.90	5.21	1.44	4.88	1.48

Table 3: Rates, mean of relevancy and easiness for the different scaling strategies.

5.2.3 Translations

Three strategies appear from the observations made on translations:

- ***Push***: the user begins his gesture outside the cube and draws it towards the object as if he was pushing it.
- ***Without object referent***: the user performs a gesture outside the cube and the trajectory is straight and collinear to the TD, without intersecting the cube.
- ***Grab***: the user picks a point of the cube and follows the trajectory of the translation. 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.

T3 is included in the *Push* strategy, T1 and T5 in the *GrabLateral* one, T4 in the *Without object 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 (see Section 4.2.3). This gesture was performed to the bottom for translating the cube to the the top (front view). 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 4 shows statistical results about these strategies. The user relevancy depends on the strategy (χ^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*, which represent gestures that are not included in one of the strategies described above, is less relevant than the *GrabAndPush* one (pF<0.01) and the *Push* one (χ^2=8.23, p<0.005). Similarly to rotations, the *Grab* strategies are more used than other strategies and, as supposed before, this may be because they rely on mouse interaction habits. Interestingly, the scores of relevancy and easiness for the *Push* strategy are higher than for the *Grab* strategies. We suggest that this observation could be due to the fact that the *Grab* strategies require more precise selections.

Strategies	%	Relevancy (QT1)		Easiness (QT2)	
		mean	SD	mean	SD
Push	15.60	6.24	1.04	6.43	0.68
W/o object R.*	4.96	4.86	1.07	5.29	0.49
GrabLateral	40.42	5.92	0.96	6.17	0.82
GrabAndPull	3.55	5.80	1.30	6.00	1.00
GrabAndPush	24.82	6.11	0.76	6.17	0.86
Others	10.56	5.07	1.83	5.67	1.40

Table 4: Rates, mean of relevancy and easiness for the different translation strategies. (* Without object referent)

5.3 Strategy distribution per user

Several strategies are used for each transformation type (rotations, scaling or translations). In this section, we investigate from the user point of view how many strategies are applied for each transformation type. Figure 5 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. However, the use is less than 100% for a significant number of cases. It means that participants have used several strategies for each transformation type. We also found that each strategy has been 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.

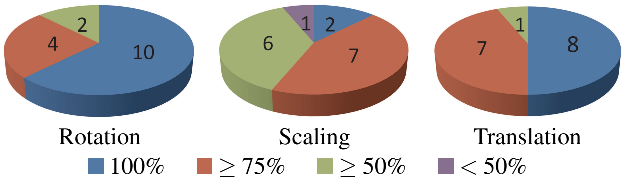


Figure 5: Use of the predominant participant strategy (16 participants). 100% means that a participant uses always the same strategy, $\geq$ 75% means that he uses a strategy with at least 75% of his gestures, and so on. The numbers in the graphs indicate the user number for each category.

6 Common behaviors

6.1 Single parameter analysis

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

Trajectories for each transformation type A key point for an interaction technique is to detect *a-priori* the transformation the user intends 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 trajectory of the gesture. Most of rotation gestures are curved (72.54%), whereas translation gestures are straight (98.92%). However 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, the detection of what the user wants to do at the beginning of the gesture in real time 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 it *a-posteriori* if the finger 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 cases where rotation circle and 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 partipants have performed such kind of gestures).

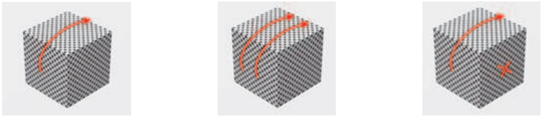
6.2 User-defined trajectory set

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 trajectory of the gestures can be used to this purpose. Note that if several fingers have the same trajectory, only one of these fingers is considered (see Figure 6).

Common behavior The agreement of trajectories for each video is evaluated using the formula introduced by Wobbrock *et al.* [13], which is:

$$\frac{\sum_{r\in R}\sum_{P_i\subseteq P_r}\left(\frac{|P_i|}{|P_r|}\right)^2}{|R|}$$

In our analysis, r is a transformation in the set of all transformations R, P_r is the set of proposed trajectories for the transformation r, and P_i 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 $(6/16)^2 + (10/16)^2 = 0.53$. Note that $|R|$ is equal to 1 because the transformations are analyzed one by one. For the whole set of gestures, the agreement about trajectories is high: 20 trajectories have an agreement greater than 0.7 (mean = 0.95, SD = 0.09), six trajectories between 0.5 and 0.7 (mean = 0.56, SD = 0.07) and one trajectory agreement is equal to 0.4 and correspond to *translation in depth* for front viewed cube. These results indicate that an interaction technique based on trajectories could be designed. One major advantage of using such an interaction technique is that it takes into account most of the gestures users draw *naturally*, without the need to learn an arbitrary language.



(a) (b) (c)

Figure 6: Definition of trajectories taken into account. The second (b) and the third (c) 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) as the (a) trajectory.

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 are, the less conflicts in trajectories there are (χ^2=58.5, p<0.0001). The maximal coverage of trajectory set without any ambiguity is equal to 83.57% of the whole gestures.

7 Discussion and conclusion

7.1 User understanding

Many questions about 3D interaction with touchscreens have been little explored. 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 set of gestures? Is this set of gestures 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 rotate it, 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), which confirms Wobbrock *et al.* observations [14]. Furthermore, users tend to follow a common behavior in a given situation (Q1) and a set of conflict-free trajectories defines a wide majority of gestures (Q6).

7.2 Implications for design

We studied the gestures performed on touchscreens by novice users in 3D 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. According to these results, a wide majority of gestures 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 a maximum of gestures is 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. The related gestures have been rated as the most easy to perform.

For a vast use and to support several strategies, favor interaction techniques based on gesture trajectory analysis. As described in Section 6.2, 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.

7.3 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 are performed according to one direction per axis only, for minimizing the duration of the study and, consequently, the fatigue of the user. However, 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 choice of the strategy. We did not manage to validate this hypothesis in this study. An new analysis dedicated to the influence of the point of view would definitively extend our results.

In our study, basic transformation types on a basic shape and with local axes have been studied to understand user gestures, in a first step. It could 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. In the same manner, it could be interesting to extend this study to physical properties and to the size of the object (*e.g.*, to know if more fingers are used when the virtual object looks heavier or bigger). The influence of the object position on the choice of the strategies could also be explored (*e.g.*, 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). 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. A comparison between gestures performed with a mouse or other devices on the one hand, and with a touchscreen on the other hand, could bring some interesting findings about the influence of the device in the choice of the trajectory. Finally, some ambiguities have been detected in the analysis of the trajectories. 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.

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

Acknowledgements

We wish to thank the reviewers, Jérôme Baril and Romain Pacanowski 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.

References

[1] A. Cohé, F. Dècle, and M. Hachet. tbox: a 3d transformation widget designed for touch-screens. In *Proceedings of CHI'11*, pages 3005–3008.

[2] K. Dandekar, B. I. Raju, and M. a. Srinivasan. 3-D Finite-Element Models of Human and Monkey Fingertips to Investigate the Mechanics of Tactile Sense. *Journal of Biomechanical Engineering (2003)*, 125(5):682.

[3] J. Epps, S. Lichman, and M. Wu. A study of hand shape use in tabletop gesture interaction. In *CHI EA'06*, pages 748–753.

[4] M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *Proceedings of CHI'07*, pages 1147–1156.

[5] M. Hancock, T. ten Cate, and S. Carpendale. Sticky tools: full 6dof force-based interaction for multi-touch tables. In *Proceedings of ITS'09*, pages 133–140.

[6] K. Kin, T. Miller, B. Bollensdorff, T. DeRose, B. Hartmann, and M. Agrawala. Eden: a professional multitouch tool for constructing virtual organic environments. In *Proceedings of CHI'11*, pages 1343–1352.

[7] H. M. K. Koskinen, J. O. Laarni, and P. M. Honkamaa. Hands-on the process control: users preferences and associations on hand movements. In *CHI EA'08*, pages 3063–3068.

[8] A. Martinet, G. Casiez, and L. Grisoni. The effect of dof separation in 3d manipulation tasks with multi-touch displays. In *Proceedings of VRST'10*, pages 111–118.

[9] A. Martinet, G. Casiez, and L. Grisoni. The Design and Evaluation of 3D Positioning Techniques for Multi-touch Displays. In *Proceedings of 3DUI'10*, pages 115–118.

[10] C. North, T. Dwyer, B. Lee, D. Fisher, P. Isenberg, G. Robertson, and K. Inkpen. Understanding multi-touch manipulation for surface computing. In *Proceedings of INTERACT'09*, pages 236–249.

[11] J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of UIST'09*, pages 69–78.

[12] A. Roudaut, E. Lecolinet, and Y. Guiard. Microrolls: expanding touch-screen input vocabulary by distinguishing rolls vs. slides of the thumb. In *Proceedings of CHI'09*, pages 927–936.

[13] J. O. Wobbrock, H. H. Aung, B. Rothrock, and B. A. Myers. Maximizing the guessability of symbolic input. In *CHI EA'05*, pages 1869–1872.

[14] J. O. Wobbrock, M. R. Morris, and A. D. Wilson. User-defined gestures for surface computing. In *Proceedings of CHI'09*, pages 1083–1092.

[15] M. Wu, C. Shen, K. Ryall, C. Forlines, and R. Balakrishnan. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (2006)*, pages 185–192.