

Understanding Hand Degrees of Freedom and Natural Gestures for 3D Interaction on Tabletop

Rémi Brouet^{1,2}, Renaud Blanch¹, and Marie-Paule Cani²

¹ Grenoble Université LIG, ² Grenoble Université LJK/INRIA
{remi.brouet,marie-paule.cani}@inria.fr,
renaud.blanch@imag.fr

Abstract. Interactively creating and editing 3D content requires the manipulation of many degrees of freedom (DoF). For instance, docking a virtual object involves 6 DoF (position and orientation). Multi-touch surfaces are good candidates as input devices for those interactions: they provide a direct manipulation where each finger contact on the table controls 2 DoF. This leads to a theoretical upper bound of 10 DoF for a single-handed interaction. With a new hand parameterization, we investigate the number of DoF that one hand can effectively control on a multi-touch surface. A first experiment shows that the dominant hand is able to perform movements that can be parameterized by 4 to 6 DoF, and no more (i.e., at most 3 fingers can be controlled independently). Through another experiment, we analyze how gestures and tasks are associated, which enable us to discover some principles for designing 3D interactions on tabletop.

Keywords: 3D manipulation, multi-touch interaction, tabletop interaction, gesture-based interaction.

1 Introduction

The interactions used to create or edit 3D content need to control simultaneously a large number of degrees of freedom (DoF). For instance, the classical docking task (i.e., defining the position and orientation of an object) requires the control of 6 DoF. The recent rise of tabletop devices seems promising for enabling such 3D interactions. Indeed, those devices have a number of desirable properties: first, despite the mismatch between the 2D nature of the input and the 3D nature of the virtual objects to be manipulated, tabletop interaction is closer to traditional shape design tools (such as pencil and paper, or modeling clay on a support table) than many 3D input devices, requiring to be hold in mid-air. Resting on a horizontal table induces less fatigue, allowing longer periods of activity. It also enables more precise gestures. Lastly, with the advent of multi-touch devices, the number of DoF that can be simultaneously controlled on a tabletop device is high: since each fingertip specifies a 2D position, the use of a single hand theoretically allows the control of 5 fingers \times 2D = 10 DoF.

This value of 10 DoF is clearly an upper bound of the actual number of DoF that a user can simultaneously manipulate with a single hand. Several evidences show that the actual number is lower: so far, no multi-touch interaction uses the positions of the

five fingers of a hand to control 10 parameters of the object being manipulated. Our common sense tells us that our fingers are not totally independent, since they are linked by the hand, and moreover that even for movements that would be physically doable, we can hardly control each finger independently.

To analyze gestures and DoF, using a new hand parameterization, we successfully decomposed gestures into elementary motion phases, such as translation, rotation and scaling phases. Such phase analysis method permits us to investigate fundamental behaviors of hands and gestures.

A first goal of this paper is to evaluate the upper bound of the number of DoF that can be simultaneously controlled by a hand on a multi-touch device. This is done through an experiment that confirms and refines what our common sense, as well as what the corpus of current multi-touch interaction techniques tell us: the number of DoF of the hand on a surface is between 4 and 6.

A second goal of the paper is to study how those DoF can be mapped to actual 3D manipulations, i.e., which interactions are the most efficient to exploit those DoF. Despite interaction with 3D content on tabletops is not “natural”, in the sense that there is no consensus among participants on how nontrivial 3D manipulations should be performed through 2D gestures, through another experiment, we discover some principles for designing 3D interactions on tabletop, which enable us to disambiguate 3D content manipulations. Possible manipulations correspond to navigation tasks (when the point of view is manipulated), object positioning tasks (i.e., object translation, rotation or scaling) and object deformation tasks (i.e., stretching, compressing or bending some part of an object).

Finally, to compare and validate our research, we investigate how the new phase analysis method fits with the other recent results on multi-touch devices.

2 Related Work

The first manipulation tool humans ever use is their hand, which enables them to touch, grab, pinch, move, or rotate many objects. Thanks to multi-touch devices, these abilities are nearly extended to the digital world.

Before creating a 3D user interface for multi-touch device, understanding the hand gesture is mandatory. Two aspects need to be studied: the hand gestures themselves, and the mapping between these gestures and tasks.

2.1 Hand/Finger Dependencies

Hand gesture analysis is a broad topic connected to many research fields. Every area we have explored notes dependences between fingers while performing a movement or a task. From a mechanical point of view, the hand consists of twenty-seven DoF, although biologically speaking, fingers are linked together by tendons and nerves and so on [1]. Neuroscientists note that a majority of hand movements can be described using two principal components [2]. Martin et al. observe dependence between fingers during voluntary and involuntary finger force change [3].

2.2 Multi-touch Interactions

The manipulation of 3D contents on tabletop is a recent research topic. Hancock et al. compared different techniques to manipulate 3D objects with one, two or three fingers [4]. They extended the RNT (for Rotation ‘N’ Translation) algorithm [5], and showed that, using spatial modes, one touch input is sufficient to control 5 DoF, while three touch inputs enables the decoupling of interactions and thus becomes more user-friendly. Martinet et al. described techniques to translate 3D objects along the depth axis using a finger of the non-dominant hand together with a unmoving dominant hand [6].

Those works are just two examples of the many 3D user interfaces using tabletop (e.g., [7–9]). A common characteristic of those interactions lies in the limited number of fingers used to manipulate the objects. Indeed, three fingers by hand are used to interact with the virtual environment for the most complex tasks, and the use of the five fingers only occur if the gesture performed is simple (a global translation and/or rotation involving the whole hand).

This rule even holds for interaction techniques designed for tasks more abstract than the manipulation of 3D contents like contextual menus that visualize information [10], or that enable the selection of tools or the switching of modes for manipulating objects [11, 12]. Again, all these interactions, while designed specifically for multi-touch devices, use at most three fingers by hand. Bailly et al.’s works about finger-count menus is a rare exception to this general pattern [13, 14]. Indeed, the number of finger corresponds to the number of the selected field in the menu.

2.3 Hand Gestures Analysis

In the context of multi-touch devices, hand gestures have been analyzed in conjunction with their mapping to particular tasks. Wobbrock et al. studied the “naturalness” of such mapping by letting users define gestures for a given set of tasks [15, 16]. Cohé et al. focused their analysis on object positioning tasks, and demonstrate the importance of finger starting points and of hand forms and trajectories [17].

In contrast, gestures are analyzed by phase analysis techniques. Nacenta et al. studied gestures during object positioning tasks, and discover that an order of manipulation exists [18]. One goal of this paper is to discover principles in order to develop 3D interactions based on phase analysis techniques.

3 Understanding Hand DoF on a Surface

To get a better understanding of possible hand gestures when the fingertips are constrained to remain on a table, we ran a first experiment that does not involve any 3D task. Since our goal was to estimate the number of DoF a user is able to simultaneously control with a single hand, we asked participants to use their dominant hand to perform a number of specific gestures.

3.1 Tasks

The gesture is specified by a starting position and an ending position. Those positions consist of five circles, each circle (resp. labeled with 1, 2, etc.) representing the position of a finger (resp. the thumb, the forefinger, etc.), as depicted on Fig. 1.a. Once a finger is correctly positioned, the corresponding circle turns green. Once all fingers are correctly positioned, the circles vanish, and the ending position appears. Then, the participant has to move his/her fingers to match the ending position, while keeping all fingers in contact with the surface. She/he can take as much time as she wants to perform each gesture.

The experiment was composed of thirty-seven trials. Those thirty-seven gestures are designed to be of various complexities: the simpler ones only involve movements of the whole hand, while the more complex ones involve the combinations of both hand movements and individual uncorrelated finger movements. Our set of gestures set was designed by testing in a preliminary study a comprehensive combination of elementary movements, and by discarding those that were too difficult to perform.

For the first ten trials, an animation between the starting and the ending position was shown to the user prior the trial, whereas for the other trials, no path was suggested. The participants were not asked to follow the suggestion, and its presence had no noticeable effect on the results we report below.

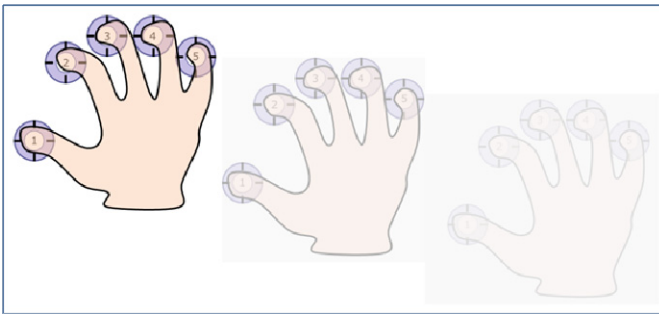


Fig. 1. To analyze hand gestures, we asked users to move their fingers from specific initial positions to specific final positions

3.2 Apparatus and Participants

This experiment was conducted on a 22'' multi-touch display by 3M (473×296 mm, 1680×1050 pixels, 60 Hz, 90 DPI). The software was based on the QT and Ogre3D libraries.

31 participants, composed of 8 women and 23 men, were tested. Average age was 30 (min. 22, max 49). All participants had normal or corrected to normal vision. For left-handed participants, the experiment was mirrored. Participant's background was variable, and not only computer scientist background. Participants' experience with 3D applications, and tactile devices was variable, but this was not an issue, as the goal of the experiment was to get some understanding of fundamental physical behavior.

3.3 New Parameterization for Hand Analysis

During each trial, the trajectories of the tip of the fingers were recorded. To analyze gestures, we define the following parameterization of the hand: we use the position of the thumb as the origin of a local frame, in order to simplify the decomposition into phases. The first axis of the frame is given by the thumb/forefinger direction of the starting position. Therefore, the hand position is given by the local frame (2 DoF for the position of the origin), and by the position of each finger in this frame (0 DoF for the thumb as it is always at the origin, 2 DoF –distance and angle– for the other fingers).

The position of each finger in the local frame is parameterized by a couple (R_i, S_i) –for rotation and scale– where R_i is the angle defined by the finger of the local frame (i.e., the angle between the thumb/forefinger direction at the starting position and the thumb/finger direction at the current position); and S_i is the ratio between the current distance to the thumb of the finger and its distance to the thumb at the starting position (Fig. 2). With these definitions, a simple translation of the hand keeps the couple (R_i, S_i) constant (only the origin of the local frame changes); a rotation of the hand changes all the R_i by the same amount but does not impact the S_i . In contrast, a pinch gesture will only impact the S_i , making them decrease from 1 (fingers at the same distance from the thumb than while resting in the starting position) to a value smaller than 1 (fingers closer to the thumb).

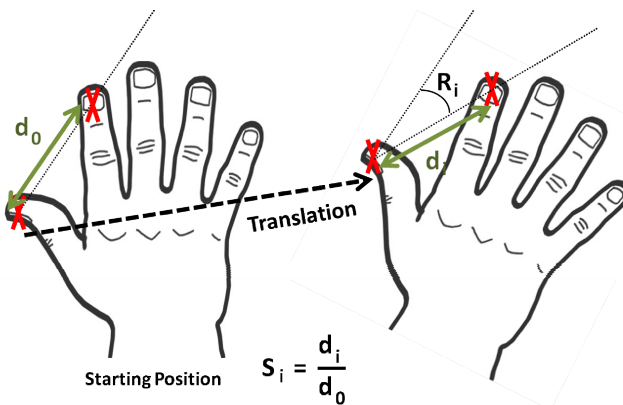


Fig. 2. Hand parameterization: definition of R_i and S_i

3.4 Results

A first look at the traces produced by participants' fingers confirms an intuitive hypothesis: hand gestures on a table can be decomposed into global motion phases (Fig. 3) and some local motion phases.

3.4.1 Global Gestures

The global part consists of the position of the hand (hand translation), its orientation (hand rotation), and how much it is opened (hand scaling). We quantify those parts using the hand translation (T) as the position of the origin of the local frame (i.e., of the thumb); and the hand rotation (R) (resp. scaling (S)) as a weighted barycenter of the R_i (resp. S_i). The weights are chosen to reduce the impact of a finger that is far from the others (i.e., to provide a kind of continuous median value), e.g., for R:

$$R = \frac{\sum_i w_i R_i}{\sum_i w_i} \text{ with } w_i = 1 + \sum_{j,k \neq i} (R_j - R_k) \quad (1)$$

We then define phases as periods of time during which a significant variation occurs for those variables, i.e., their first derivative is above a threshold (i.e., threshold are respectively 0.005, 0.5, and 0.001 for translation, rotation and scaling). Fig. 3 shows the variation of T, R and S while performing a gesture (top), and the corresponding phases (bottom). The pattern formed by this example is typical of what can be observed: there is a single phase for the translation, while the rotation and scaling are achieved during several phases (typically less phases are needed for R than for S). The different phases start roughly at the same time but end in this order: first T, then R, and then S. This pattern is similar to the one observed by Nacenta et al. [18], since what they call “period of maximum activity” are the second phase for R and the second or third phase for S.

To further validate this order of manipulations, we can look at the number of phases needed to validate the trials. Fig. 4 summarizes those results: for more than 93% of the cases, users need a single translation phase to correctly position their hand; while a correct rotation is achieved within a single phase for 68% of the trials and a correct scale for 35% of the trials.

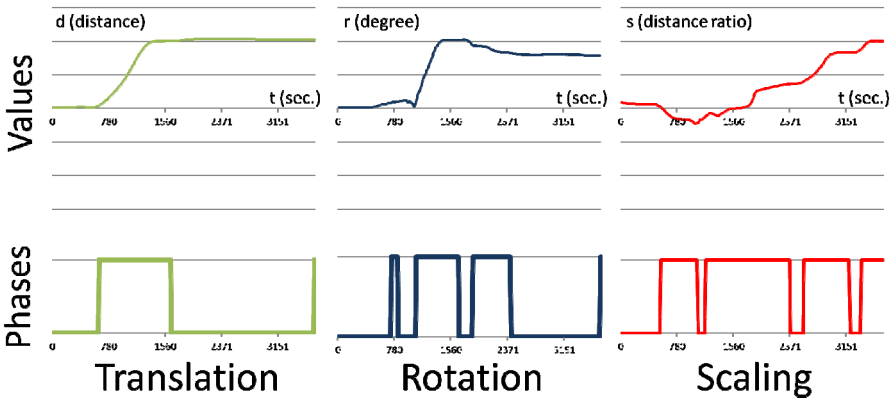


Fig. 3. Global variations (top), and corresponding phases (bottom), during a gesture: variations of translation, rotation and scaling

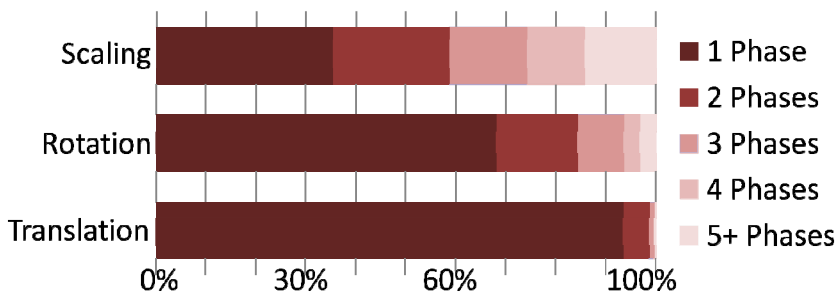


Fig. 4. Percentage of tasks where 1, 2, 3, 4, 5 or more phases are required among all tasks and participants for translation (T), rotation (R) and scaling (S)

Thus we think that hand gestures can be decomposed into sub-parts that have different degrees of stability: from the most stable motion (global translation) to the less stable motion (one finger motion). For instance (Fig. 5), the global translation is the easiest to get right (1 phase only), without any interference afterwards. On the contrary, global translation could induce interferences on rotation (first rotation phase), before that the major rotation motion is performed (second rotation phase). As translation and rotation are simultaneously performed, sometimes rotation motion has to be corrected (third phase).

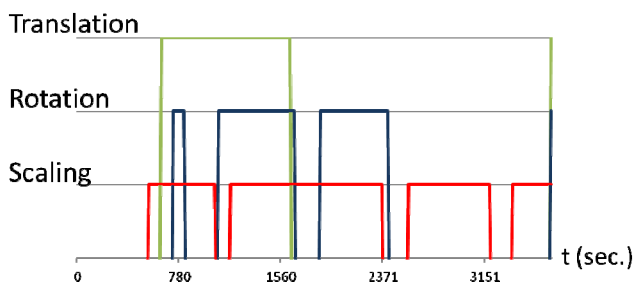


Fig. 5. Phases superposition for the Fig. 3 example

3.4.2 Local Gestures

The local parts of gesture are the components of individual finger movements that are not explained by the global T, R and S described above. A first look at the data shows that those local parts are mainly movements performed by the middle, ring and little fingers. To get a better understanding of those movements, we concentrate our analysis on the trials in which users had to perform movements involving only a subset of those fingers, and in which those movements was the same for the fingers involved.

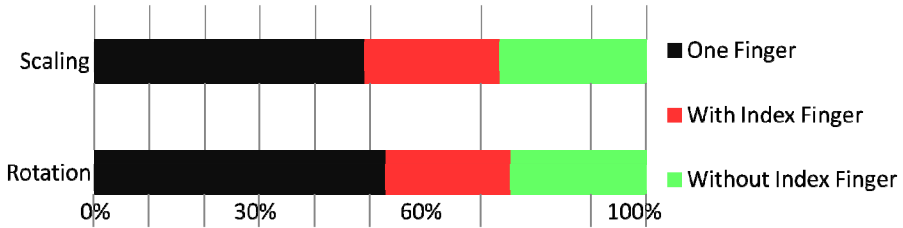


Fig. 6. Average percentage of time spent for moving a single finger, more than other fingers including/not including the index finger for tasks involving the motion of one or more fingers among the last three fingers only

For those tasks, Fig. 6 shows the proportion of time spends moving a single finger ($\approx 50\%$ of the time), the time spends moving more than one finger, including and excluding the index finger ($\approx 25\%$ each). We can note that those proportions are roughly the same whether the participants are asked to perform a scaling task, i.e., to control the S_i (top) or a rotation task, i.e., to control the R_i (bottom). It is also interesting that the movement of one (or more) of the last three fingers involves the motion of the index finger despite that in those tasks the index finger was not supposed to move. This shows how it is difficult for users to control the three last fingers simultaneously and independently. The interdependence between those fingers is consistent with the study conducted by Martin et al. [3].

To further investigate the interdependencies among the last three fingers, we split the trials into three groups, depending on the number of fingers the users have to move among the middle, ring and little fingers. Fig. 7 shows for each group (vertically: 1F, 2F, 3F), the relative time spent moving 1, 2 or 3 of those fingers. It is interesting to note that even when asked to move a single finger (1F), the participants spend more than 30% of their time moving two or more fingers. On the other hand, when participants have to perform the same motion for the last three fingers (3F), only one third of the time is used to move the fingers together, while $\approx 40\%$ of the time the fingers are moved individually.

This confirms that the three last fingers cannot be used to control something independently of the index finger, even if they are used together as a whole. Such dependencies induce difficulties for users to efficiently control the hand 10 DoF, and decrease this upper bound around 4 or 6 DoF (two or three independent fingers).

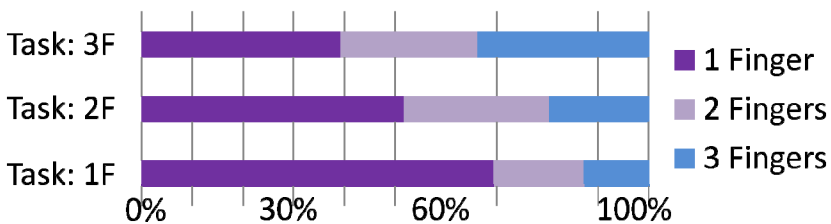


Fig. 7. Average percentage of time spent for moving 1, 2 or 3 fingers among the last three fingers, when the user is asked to move 1, 2 or 3 of them (1F, 2F, 3F)

4 Mapping Gestures and 3D manipulation

We ran a second experiment to understand the most natural mapping between 2D gestures and 3D tasks. Recent researches have focused either on navigation tasks (e.g., [4, 6]) or object positioning tasks (e.g., [7, 15]). Mixing both kinds of task increases the number of possible mapping. Therefore, one of our goals was to discover if the implicit information included in an interaction could be used to automatically switch between interaction modes, rather than having to provide explicit widgets for mode selection

4.1 Tasks

The participant observes an animation of the desired task on the first part of the screen (Fig. 8a, b) (left)), and then he/she performs a gesture of their choice to perform this task (right). The experiment was composed of thirty-six trials, divided into three classes: eleven navigation tasks, nine object positioning tasks and sixteen object deformations tasks. For navigation or object positioning tasks, the scene was composed of two cubes, a grid, and a background picture (Fig. 8a). For object deformations, only the grid and 3D object were shown (Fig. 8b).

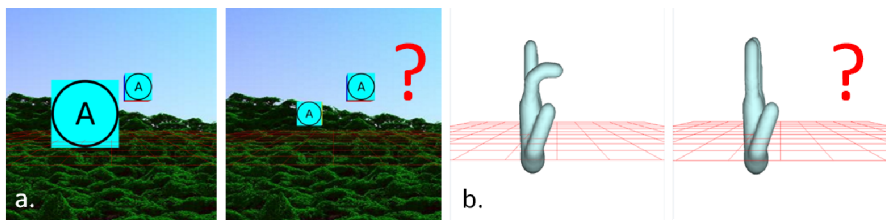


Fig. 8. a) Example of setting for discovering fundamental behavior for navigation / object positioning tasks. An animation is shown on one screen (left), while users perform gesture on second screen (right). b) Similar setting for object deformation tasks.

4.2 Hand Phase Analysis

The analysis process performed for the first experiment was reproduced with little differences. However, we had to adapt the hand parameterization to the number of fingers in contact with the table. Contrary to the first experiment, where each finger could be identified by the starting position, all interactions did not always involve the five fingers (e.g., the thumb was not always used).

The first experiment demonstrated that the thumb is usually the most stable finger (this was our reason for using it as origin of the local frame). Therefore, we assumed the thumb to be the one that was moving the less (this assumption can be wrong when the gesture is a translation, but this is non-issue, since all the fingers are being moved the same way in this case). The other fingers do not need to be distinguished. We also had to perform two distinct phase analyses, one for each hand, to interpret the gestures.

4.3 Results

4.3.1 Hands/Fingers Uses

To deeper investigate the efficient DoF a hand can control, we first observe that only three participants used more than 3 fingers by hand. Those cases mostly involved navigation tasks. In more details, when participants involved more than 3 fingers to manipulate the 3D contents, the principal phase of their interaction corresponds to translation phase (i.e., the most global motion). On average, fewer fingers by hand are used to handle objects than to navigate (Table 1). The difference between numbers can be explained by the use of the second hand. Further explanations are developed in the next section.

Table 1. Second Experiment results

Tasks	Generality		2 nd Hand Data			Phase Analysis			Best Gesture 1 st /2 nd
	Dist. to Obj.	Avg. #Fingers	% Tasks	2 nd Hand #Fingers	Type	#T. Phase	#R. Phase	#S. Phase	
Navigation									
Translation /xy	1.5	2.5	03	1.0	Sym.	1.2	0.1	2.0	Tr./-
Translation /z	1.9	2.9	46	1.7	Sym.	1.5	0.1	1.0	Tr./Sym.
Rotation /xy	2.3	2.8	52	1.4	Sup.	1.1	0.3	1.5	Tr./Sup.
Zoom	1.6	3.5	54	1.8	Sym.	1.5	0.1	1.2	Tr./Sym.
Zoom to Object	1.2	3.0	40	1.6	Sym.	1.5	0.2	1.1	Tr./Sym.
Object Positioning									
Translation /xy	0.2	1.2	00	-	-	1.0	0.1	0.3	Tr./-
Translation /z	0.5	2.1	24	1.3	Sup.	1.2	0.1	0.6	Tr./-
Rotation /z	0.5	2.3	00	-	-	0.7	1.1	1.9	Rot2./-
Rotation /xy	0.7	2.3	57	1.1	Sup.	0.9	0.1	0.9	Tr./Sup.
Scaling	0.4	3.0	19	2.5	Sym.	1.4	0.4	1.0	Sca2./-
Obj. Deformation									
Extrusion	-	1.6	42	1.4	Sup.	1.1	0.1	0.2	Tr./Sup.
Bending /z	-	2.2	19	1.3	Sup.	0.2	0.9	1.2	Rot1./-
Bending /xy	-	2.3	77	1.2	Sup.	0.7	0.2	0.2	Tr./Sup.
Local Scaling	-	2.4	21	1.8	Sup.	0.9	0.2	1.3	Sca2./-
Deleting	-	1.2	11	1.5	Sup.	1.1	0.0	0.2	Tr./-
New Object	-	2.6	17	1.3	Sup.	1.4	0.2	1.4	Sca2./-
Object Selection									
Selection	0.5	1.0	-	-	-	-	-	-	*see 4.3.2

However, many users interacted using both hands. From our observations, the non-dominant hand had two main functions: a support function (Sup.) (e.g., frequently indicating the parts of the scene that should not move by keeping a still hand on them); or a symmetric function (Sym.) (e.g., doing symmetric gestures with both hands for scaling). The support function is most frequently used, specifically on

object manipulation tasks where it is used to maintain some objects or some part of the object of interest in place.

4.3.2 Modes Disambiguation

The vast majority of users (87%) performed ambiguous gestures, i.e., used similar gestures for two different tasks. This leads us to look for ways to disambiguate those gestures.

A first clue for disambiguation is the location of the fingers at the start of the gesture: the first finger is hardly put on or around an object when a navigation task is involved (distance > 1, Table 1), directly manipulating on the background image. Furthermore, the grid is sometimes manipulated to perform indirectly navigation tasks such as panning along the depth axis. On the other hand, object manipulations typically start in or nearby the object (distance < 1). Although this criterion enables us to distinguish navigation tasks from object manipulation tasks, further investigation has to be done to disambiguate object positioning from object deformations.

A second clue for disambiguation is the number of fingers used. The average number of fingers involved to navigate is about 3 while this number decreased to 2 for object positioning. Though, the non-dominant hand gives the most relevant number of fingers: 1 finger used for navigation, no finger for object positioning and 1 or more for object deformation. In a large proportion, the non-dominant hand fingers reached the border of the screen for navigation tasks when it has a support function.

Therefore, the different modes could be automatically distinguished during user interaction by mixing these two criteria: a finger-count method [13] would give the selected interface mode, while finger locations could tell to which object the interaction is to be applied, if not to the whole scene.

4.3.2 Group Selection

Another issue investigated is how a transformation could be applied to a couple of objects. The same gesture was usually performed for both objects ($\approx 75\%$ of users), each object involving one hand. But this does not scale to more than two objects, and cannot be applied to gestures requiring both hands.

Instead of simultaneously/sequentially manipulating the different 3D contents, fewer participants ($\approx 20\%$) preferred to first select the object by clicking (or double clicking) before manipulation. Only two users performed a “lasso” gesture to select object before performing the transformation. After the object selection, the gesture was performed either on one of the object, or near the barycenter of the group. This leads us to conclude that a specific widget should be created to represent the selected group.

4.3.3 Scaling Interferences

During most tasks, the participants performed scaling phases while performing their interaction. In many cases, the DoF involved by scaling was meaningless.

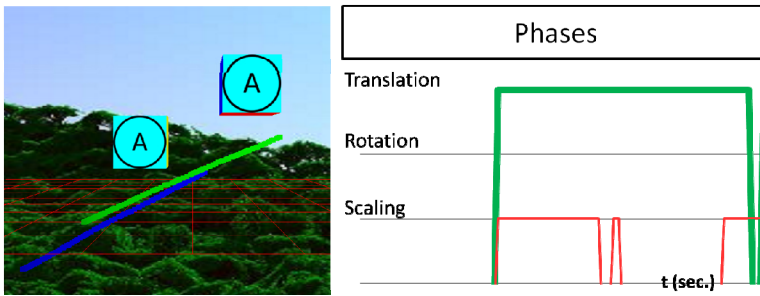


Fig. 9. Gesture for a translation task and its phases

For instance, Fig. 9 illustrates the gesture of a participant during a navigation task: a translation in the (x, y) plane. In this illustration, more than 90% of the motion was analyzed as translation phase, while short scaling phases occurred in parallel. As stressed when analyzing the first experiment, the stable and useful part of scaling motions usually takes place when the translation and rotation phases of a motion have ended. Therefore scaling phases should not be taken account when they occur concurrently to other phases, and the gesture in Fig. 9 should be interpreted as a bare translation.

4.3.4 Navigation Tasks: Zoom vs. Depth Axis Translation

Two consecutive trials were depth axis translation and zooming tasks. To distinguish the different kinds of trials, a background image was added to the 3D scene. Though, every user but two asked for the differences. Once answered, they mainly succeed to understand the shown transformation.

Moreover, we can also notice that, although they did know the difference (as they asked for it), half of the participants still performed the same gestures for both tasks.

4.3.5 Combining Different Manipulations

Some tasks of the experiment consisted in combining elementary motions – for instance, object translation and rotation. In order not to influence the participants, and leave them free to invent their own interaction mode, only a before/after animation was shown in this case. Analyzing data by phase analysis techniques enables us to easily distinguish whether users prefer to perform each “elementary” motion sequentially, or simultaneously. The results are gathered on Table 2.

Table 2. Table illustrating whether users prefer to separate the different motion (left), or not

Tasks	% Sequential motions	% Concurrent motions
Translation + z/Rotation	58%	42%
Translation + xy/Rotation	79%	21%
Translation + Scaling	61%	39%

In two third of the cases, participants preferred to decompose gestures into “elementary” ones. This is consistent with Martinet et al. work [19]. In details, performing a translation and a depth axis rotation are mainly decomposed into a translation and a rotation phases. The higher number of participants performing simultaneously these two phases (42%) are consistent to Wang and Nacenta works [18, 20], as these two phases slightly interferes each other. On the other hand, when translation is coupled with a rotation along the other axis, the phase analysis mainly identified two translation phases; the second phase corresponding to the second hand gesture: a trackball like rotation (see further details in the next section) [21].

4.3.6 Starting Finger Positions for Deformation Tasks

We already observe that starting positions of fingers is relevant in order to disambiguate navigation from object manipulation tasks. Further investigations about fingers starting positions have been performed for object deformation tasks.

When participants use their non-dominant hand, their fingers typically remain far away from part of the object that is deformed (even sometimes at the opposite side). By performing such gesture, participants keep in place the object, while she/he works on a region of interest of the object –such as designers keep in place their paper while drawing [22].

However, the dominant hand gestures are typically performed around the deformed object. For instance, on bending tasks, the thumb position corresponds to the center of rotation, and remains static, while a rotation gesture is detected by phase analysis (Table 1). Local scaling (such as stretching or compression tasks) is typically performed by a shrink gesture, where the gesture barycenter is located nearby the center of the part of the object that is being deformed.

4.3.7 Noticeable Gestural Design Pattern

As we already observed, a majority of users performed ambiguous gestures, and therefore the interface need some disambiguation between modes. On the opposite, we note that some manipulations can be linked together, enabling us to identify typical gestures and a gestural pattern for each mode.






Gestures	Translation	Rotation 1	Rotation2	Scaling 1	Scaling 2*
Phase:	Translation	Rotation	Translation + Rotation	Scaling	Translation. + Scaling
Type:					

Fig. 10. Five typical gestures for one hand interaction, identified through our experiment. Scaling 2 is difficult to identify due to scaling interferences (see section 4.3.3)

Once phases are analyzed, hand gestures on a surface can be easily classified into 5 main classes (Fig. 10). Due to scaling interferences, no gesture is identified when all three phases are detected. On the contrary, the detected gesture is Rotation 2. In Table 1 (last column), we associate each task of the user study to the corresponding typical gestures for the first hand.

The gestural pattern is summarized in Table 3. On the first hand, manipulations that transform the scene/object on the 2D screen plane mainly used one-handed gestures (e.g., translation/extrusion along x, y axis tasks are performed by one hand translation gestures). Scaling manipulations can be gathered into two possible gestures, which both represent a shrink gesture, either performed by one or two hands.

Table 3. Table grouping a set of action – either usable on navigation tasks, or object positioning/deformation tasks – and the users associated gestures

Action	Translation / xy	Translation / z	Rotation / z	Rotation / xy	Scaling / Zoom
Gestures (Phases)	Translation	?	Rotation	Translation + Support	1 or 2 handed Shrink gesture.

On the other hand, manipulations that required depth axis motions need more attention. For instance, rotation tasks are usually performed with two hands: one hand is keeping the object in place, while the second hand is “pushing” the object, like in the trackball technique [21]. Though, manipulations that correspond to a translation along depth axis are outsiders: no gesture was consistently used to perform these tasks.

Using such a gestural design pattern for all 3D multi-touch interfaces would be a real advantage, since users would need to learn the pattern only once, and would immediately be efficient with new tools.

5 Comparison with, and Application to Previous Work

5.1 Other Multi-touch Gestures Analysis: Cohé and Hachet Work

Cohé and Hachet recent research lead them to another approach of understanding gestures for manipulating 3D contents [17]. Their paper focused on object positioning tasks.

Their approach was to classify gestures using three parameters: form, initial point locations, and trajectory. They identified gestures by the number of moving/unmoving fingers (the form), their starting locations (initial point), and the kind of motion (trajectory), while exploring object translation, rotation or one axis scaling tasks.

For those tasks, while we used a different methodology, our results are largely consistent with their findings: in our case, form and trajectory parameters are considered by the phase analysis. Nearly all their classifications are coherent with the gestural pattern that we defined above. For instance, their rotation gestures (except for R3 and R8)

are identical to our rotation phase. Moreover, both papers observe that a majority of users prefer to start on or nearby the object.

The main difference is the parameters used to define the starting locations. While we only defined the neighborhood of object to distinguish between modes, they divided this parameter according to cube elements (faces, edges, corners and external). Both classifications bring their own advantages. Using cube elements to directly manipulate complex 3D content such as large triangular meshes would be meaningless. On the opposite, manipulating 3D content with 3D transformation widgets could always make use of cube like widgets, and therefore use the proposed decomposition.

5.2 Direct Interaction Techniques: 1-, 2- and 3-Touch Techniques

A first kind of 3D manipulation is interactions that are directly performed onto the objects. Hancock and Cockburn researches identify 3 techniques, based on the number of fingers used (which are extended by Martinet's works for depth axis translation) [4, 6].

Their paper is focused on the comparison between three techniques that enable users to perform translations and rotations. The first technique, involving only one-touch interactions, corresponds to an extension of the RNT algorithm [5]. By doing so, the interface can manipulate 5 DoF with a single finger. The second technique, involving two touch interactions, the first finger correspond to the RNT algorithm for translations and yaw motions, while the second finger is used to specify the remaining rotations. The last technique maps each group of motion to a specific finger – translation to the thumb, yaw rotation to the second finger and the remaining motions to the last finger.

It is noticeable that they stop their comparison up to three-finger techniques that corresponds to our effective upper bound number of fingers. They compared the three techniques in two experiments. For both tasks, they concluded that the three-touch technique was the fastest to use, while the one touch techniques was the less efficient one.

We will further focus on the differences between these methods, compared with our phase analysis method. Even though the one touch technique is the most stable gesture (as it can only provide translation phases), the technique suffers of DoF distinctions: all interactions are mapped to the same gesture. On the other hand, the three-touch technique easily decomposes translation and z-axis rotation to translation and rotation phases into the two first finger motions. Translation and rotation phases can be mainly performed at the same time, with little interference between them, so users are more efficient while performing such techniques.

Though, the last finger suffers from the same issue on two and three touch techniques. Indeed, as the roll and pitch rotation are mapped in the Cartesian frame, rotation and scaling local phase are mixed during the last finger gestures. Therefore, performing pure roll or a pitch rotation are interfering each other.

5.3 Indirect Interaction/Widget Technique: tBOX Analysis

Another kind of 3D manipulation involves a widget that acts as a proxy to the real object. 3D transformation widgets are commonly used in 3D applications. A recent

example of 3D transformation widgets for multi-touch devices is the tBOX [7]. To easily manipulate 3D objects, they are enclosed in their bounding box that is made interactive. This is an extension of the standard manipulation widget (represented by 3 arrows). The existing manipulations on objects are translation, rotation and scaling. All user gestures have to involve the cube widget – specifically the vertices, edges or faces of the cube. For instance, pushing a single edge performs rotations, while translating along edge performs translations. A shrink gesture on both sides of the tBOX widget represents a single axis scaling.

The first observation about tBOX, once analyzed into phases, is that all object manipulations are translation phases only (scaling corresponding to one hand translation and a symmetric second hand role). In terms of stability, such gestures are the most efficient, as no interferences can occur. Moreover, such widget leaves a lot of possible interactions for other manipulations (such as deformations).

To the tBOX authors mind, one goal of their interactions was to discriminate between rotation and translation. Therefore, users cannot efficiently switch between these two manipulations: they have to stop their first gesture and reach again the required edge. On the other hand, phase analysis based interface would permit to easily switch between these manipulations, maybe at the cost of stability.

6 Discussion

Theoretically, multi-touch devices offer the possibility of manipulating 3D scenes while simultaneously controlling many DoF: up to 20 actually, if the two hands were used. However, this upper bound is never reached. Because of the interferences between fingers and to their restricted motion when moved in contact with a plane, complex gestures involving all fingers are often unstable, and the time it takes to perform them would be prohibitive for an interactive use.

As shown by the second experiment, users easily invent gestures to interact with 3D content. Quite interestingly, they tend to use all fingers for global hand gestures such as translation, rotation, and scaling, although two or three fingers would be sufficient (in this case, using all fingers is easy, since there is no local hand motion to control). For more complex interaction gestures, users naturally limit themselves to one to three fingers per hand. This leads us to the following methodological rules when designing 3D interaction on a multi-touch table:

- Firstly, the number of DoF effectively controlled by the user (never more than 8 for the two hands in our experiments) is actually much smaller than the number of DoF required for navigating, plus moving and deforming objects in a 3D scene. Therefore, using an interaction system based on several interaction modes is mandatory.
- Secondly, the number of fingers actually on the device during the interaction gesture could be easily used to distinguish between simple navigation tasks, and more complex object positioning/deformation tasks: full hand interaction could be used to select and control navigation, since simple global gestures, which the user preferably performs with all fingers, are sufficient in this case. For object

manipulation/editing tasks, the interface could disambiguate the required mode by counting the number of finger on non-dominant hand.

- As noted in our experiments, the location where the gesture starts is often meaningful: users typically use it to select the object to which the action is applied. In addition to controlling object selection, the hand location at the start of the gesture could be another way of automatically selecting between navigation (if the gesture starts on the background) and object positioning (with some limitation for crowded scene, where some free background space would need to be artificially preserved for navigation).
- Global phase analysis is quite coherent for mapping gestures and tasks: gestures are easily classified. Even more, a design gestural pattern for 3D contents manipulations emerged from the experiments, which are reproduced inside each tested mode, and could be extended to any other 3D content transformation mode. Though, scaling phases should be analyzed independently, when the other gesture phases have stopped, as they can be produced as side effect of other phases.
- Lastly, using the full hand to grab groups of objects on which to apply a gesture (such as all the objects covered by finger tips, or by the convex envelop of finger tip positions) would be a further extension of this technique. However, extra gestures such as double-clicking with a finger, or circling the object to select it (as done by some of our users), would be needed to add distant objects to the group.

7 Future Work

The first goal of this paper was to understand hand gestures on a surface. The phase analysis technique we proposed provides a simple, yet consistent way to analyze and classify gestures, especially regarding global hand motion. Therefore, an interesting direction for future research would be to develop new interaction methods directly relying on such phase analysis to drive task control.

Acknowledgements. This research was partially funded by ERC advanced grant EXPRESSIVE and by the G-INP BQR Intuactive.

References

1. Wilhelmi, B.J.: Hand anatomy (2011)
2. Santello, M., Flanders, M., Soechting, J.F.: Postural hand synergies for tool use. *J. of Neuroscience* 18, 10105–10115 (1998)
3. Martin, J.R., Zatsiorsky, V.M., Latash, M.L.: Multi-finger interaction during involuntary and voluntary single finger force changes. *Exp. Brain Research* 208, 423–435 (2011)
4. Hancock, M., Carpendale, S., Cockburn, A.: Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In: *Proc. CHI 2007*, pp. 1147–1156 (2007)
5. Kruger, R., Carpendale, S., Scott, S.D., Tang, A.: Fluid integration of rotation and translation. In: *Proc. CHI 2005*, pp. 601–610 (2005)

6. Martinet, A., Casiez, G., Grisoni, L.: The design and evaluation of 3d positioning techniques for multi-touch displays. In: Proc. 3DUI 2010, pp. 115–118 (2010)
7. Cohé, A., Declé, F., Hachet, M.: tBox: A 3D Transformation Widget designed for Touchscreens. In: Proc. CHI 2011, pp. 3005–3008 (2011)
8. Grossman, T., Wigdor, D., Balakrishnan, R.: Multi-finger gestural interaction with 3d volumetric displays. In: Proc UIST 2004, pp. 61–70 (2004)
9. De Araújo, B.R., Casiez, G., Jorge, J.A.: Mockup builder: direct 3D modeling on and above the surface in a continuous interaction space. In: Proc. GI 2012, pp. 173–180 (2012)
10. Francone, J., Bailly, G., Lecolinet, E., Mandran, N., Nigay, L.: Wavelet Menus on Handheld Devices: Stacking Metaphor for Novice Mode and Eyes-Free Selection for Expert Mode. In: Proc. AVI 2010, pp. 173–180 (2010)
11. Hancock, M., Hilliges, O., Collins, C., Baur, D., Carpendale, S.: Exploring tangible and direct touch interfaces for manipulating 2D and 3D information on a digital table. In: Proc. ITS 2009, pp. 77–84 (2009)
12. Scoditti, A., Vincent, T., Coutaz, J., Blanch, R., Mandran, N.: TouchOver: decoupling positioning from selection on touch-based handheld devices. In: Proc. IHM 2011, pp. 37–40 (2011)
13. Bailly, G., Müller, J., Lecolinet, E.: Design and evaluation of finger-count interaction: Combining multitouch gestures and menus. *IJHCS* 70, 673–689 (2012)
14. Bailly, G., Demeure, A., Lecolinet, E., Nigay, L.: MultiTouch menu (MTM). In: Proc. IHM 2008, pp. 165–168 (2008)
15. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proc. CHI 2009, pp. 1083–1092 (2009)
16. Morris, M.R., Wobbrock, J.O., Wilson, A.D.: Understanding users' preferences for surface gestures. In: Proc. GI 2010, pp. 261–268 (2010)
17. Cohé, A., Hachet, M.: Understanding user gestures for manipulating 3D objects from touchscreen inputs. In: Proceedings of the 2012 Graphics Interface Conference, pp. 157–164 (2012)
18. Nacenta, M.A., Baudisch, P., Benko, H., Wilson, A.: Separability of spatial manipulations in multi-touch interfaces. In: Proc. GI 2009, pp. 175–182 (2009)
19. Martinet, A., Casiez, G., Grisoni, L.: Integrality and separability of multitouch interaction techniques in 3D manipulation tasks. *IEEE Transactions on Visualization and Computer Graphics* 18, 369–380 (2012)
20. Wang, Y., MacKenzie, C.L., Summers, V.A., Booth, K.S., et al.: The structure of object transportation and orientation in human-computer interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 312–319 (1998)
21. Chen, M., Mountford, S.J., Sellen, A.: A study in interactive 3-D rotation using 2-D control devices. In: *ACM SIGGRAPH Computer Graphics*, pp. 121–129 (1988)
22. Guiard, Y.: Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of Motor Behavior* 19, 486–517 (1987)