

# The Study of the Full Cycle of Gesture Interaction, The Continuum between 2D and 3D

Mohamed-Ikbel Boulabiar<sup>1</sup>, Gilles Coppin<sup>1</sup>, and Franck Poirier<sup>2</sup>

<sup>1</sup> Lab-STICC, Telecom Bretagne, France

{mohamed.boulabiar, gilles.coppin}@telecom-bretagne.eu

<sup>2</sup> Lab-STICC, University of Bretagne-Sud, France  
franck.poirier@univ-ubs.fr

**Abstract.** The goal of HCI researchers is to make interaction with computer interfaces simpler, efficient and more natural. In a context of object manipulation, we think that reaching this goal requires the ability to predict and recognize how humans grasp then manipulate objects. This is based on studies explaining human vision, reach, grasp taxonomies and manipulations. In this paper, we study the full cycle of gesture interaction using different points of view, then attempt to organize them using Norman's theory of Human Action, we link the psychology of object sensing to HCI goals and propose a simplification of gestures classes into four principal families. Our simplification of gestures classes still allow the expression of more detailed subclasses differentiated by the gesture properties.

**Keywords:** Gesture, 3D, Interaction, Hand, Grasping.

## 1 Introduction

From the birth, human beings tend to discover their environment using all their senses. They discover the existence of things by sight, then tend to touch, grasp and manipulate objects before using these recognized items or tools to accomplish other tasks [35]. Our study focuses on the whole cycle of an interaction [27,26] starting from observation, through grasping and focusing in manipulation and specifically direct manipulation of objects. We attempt to link many research areas for the quest of natural and direct interaction as expressed by Beaudouin-Lafon [2].

This research around natural interaction in 2D and 3D space is a high priority because interfaces and visualization techniques have evolved from command line to graphical user interface (GUI) with multidimensional graphical elements, but the interaction are still almost the same. The presence of mid-cost 3D stereoscopic displays, using special glasses, provides an immersion of 3D objects below and above the rendering surface. Some researchers worked on controlling such rendering with 2D multi-touch input [34] but this is still very limited and a real 3D spatial manipulation is more appropriate for direct interaction. In order to reflect real world manipulations with objects, many frameworks like [11]

or [15] have abstracted the interaction between the hand and objects to those emulated by physics simulation engines. Even if these ways of interaction with objects provide a sense of reality, they still provide neither information about the psychology of grasping and manipulation, nor a prediction of the upcoming position of the hand and fingers. Hence comes the importance of psychological research to fill the gap between the goals of HCI among naturalness, efficiency, discoverability and the current state of studies.

One of the main foci of Human-Computer Interaction researchers is to make interaction easier and more intuitive allowing a larger spectrum of people to better use systems. This is called making interaction natural in the meaning of reducing the need to remember complex operations. Actions should be easily discoverable and the system may be learned through exploration. From the other side, there have been many studies on human dexterity and hand grasping taxonomies. Napier in [24] has proposed the classification of human grasp into two main categories: power grasp and precise grasp. According to this simple classification, we think that we can link studies on grasp and the gesture generation cycle with the objective of HCI research and start a new conquest for a more natural interaction.

Our contributions are : 1. Attempt to link between multi-fields of research around gesture, 2. Proposition of a new simplified taxonomy of gestures, 3. Rising the problematic of gestures prediction with virtual objects.

This paper proposes a representation of previous work on neuropsychology, grasp and gesture from different points of view and by using Norman model of human interaction to organize the separate areas of research, then explaining the limits of stopping the studies at grasping and not continuing to manipulation. In the next chapter we discuss manipulation of objects and we propose a simplified taxonomy.

## 2 Overview of the Related Work Around the Hand and Gesture

### 2.1 The Evolution of the Hand with Tool Use

When the first primates have left the trees, they started freeing their hands for new uses. As Napier [23] refers to, the use of hands in hominids evolved from self feeding. In contrast to most animals which use their both hands, hominids and some great apes are able to use one hand to grasp objects thanks to the thumb. It was thought that humans are the only creatures with the ability to create tools instead of just using them, they were even called “Homo Faber”. Even if it cited in Napier’s book, it was discovered recently that apes can use and even create tools [32]. This evolution of the ability to grasp, then the use and creation of tools in primates, diverged just in a small fraction that made the anatomy of human hand with more dexterity only for precise gestures manipulations like playing musical instruments [19].

## 2.2 Gestures and the Speech in the Brain

The recent man and the great apes like the chimpanzees can use hand gestures for communication [31]. David McNeill has emphasized on the relation between the gesture and the speech. In his book [20] he made an analysis of gestures relatively to speech. The taxonomy he proposed categorize gesticulation, speech-linked gestures, emblems and pantomime. The taxonomy of gestures vis-a-vis speech has its background as their are both proceeded by the same neural system in the brain [40].

## 2.3 Motor and Vision Brain Pathways

Manipulating an object requires detecting its presence mainly using the eyes, then planning an action. An experiment run by Aglioti et al. [1] demonstrated that the visual illusions have not impacted the motor action. This experiment has proven that the perception and the action are two separate paths that do not interfere with each others [5]. Another work has suggested that even for a hand action to occur, there are two pathways, one for moving the fingers and another for transporting the hand itself [29]. More detailed studies about hand movement and the physiology of grasping are made by Nowak et al. [28] studying subjects with disorders in the somatosensory system, or with the Parkinson's diseases among many others.

## 2.4 Proprioception, Exteroception and Manipulation Area

The proprioception is the individual perception of himself, and the self sense of the parts of the body. [38]. This knowledge is acquired during the first years of the individual existence and decline by aging. The proprioceptive kinesthetic sensory system intervene in controlling and correcting limb movements during a reach movement.

The importance of proprioception in a 3D gesture manipulation has been studied by Mine et al. [21]. The lack of feedback makes an interaction very difficult and should be compensated by the proprioceptive capacity. Memorizing real world positions by the exteroception, allow an efficient acces and reach to the objects. In reaching study, it is easier to remember a position relative to ones own hand more accurately than a position fixed in a virtual space.

## 2.5 Troubles in gestures Choices and Manipulation: Gestemes and Kinemes

The planning of manipulation starts with knowledge acquired in proprioception, which is the self sense of the parts of the body. It's a learning process which starts from the childhood and keeps enhancing through the age. In the grasping process, proprioception is primordial to orient the palm and reshape the hand [17].

The classification of Signoret and North cited in [33] considers the two distinct functional levels in gestures after a study on people having apraxia problems,

and who fail in choosing either the right gesture for a specified object, or in performing the gesture through the right list of kinemes. These levels explained more in Bertranne thesis [4], are the goal of the gesture and the motor realization. Gesture production needs in the same time the choice of the right gesteme, then the sequence of kinemes performed with the hand. During the experiments, patients were asked either to just mimic a just performed gesture to evaluate the ideomotor apraxia, or to perform a named gesture to evaluate the ideational apraxia.

## **2.6 Grasp, Hand Shape and Fingers Configuration**

The human dexterity in grasping is efficient but very difficult to mimic. Researchers who want to build robots capable of grasping and moving objects focus on human grasping creating taxonomies which would be possible to mimic on robots [6]. The need for a robot capable of performing a multitude of tasks, Cutkosky [8] and Feix et al. [10] started by studying human grasp selection. These robots needs to grasp objects in order to manipulate them which include moving or rotating. Even if the goal is for robots, the main element in the process is the human, his/her hand and its dexterous manipulation [36] .

## **2.7 The Search for the Best Gesture Interaction in HCI**

Many Human-Computer Interaction researchers have been working on tracking the best gestures to be used for a system. In their research, like the one conducted Wobbrock et al. and Gustafson et al. [39,13], they asked subjects during an experiment to perform what they thought would be the best gesture to achieve a specific task. During so, we may argue that the methods used in these studies are more statistical analyses than a proper reverse engineering of the human behaviors. Other studies focuses on finding newer methods of interaction or new gestures like Nacenta et al. work [22], or by the introduction of different sensors for the human manipulation [14].

## **2.8 Simulating a Rich Interaction**

Physics simulation libraries are used to add a feeling of real interaction with objects like in the work by Frohlich et al. [11]. Given the properties of objects, Newtonian forces are computed to move, distort or bounce an object according on its type and how a user is touching it. Systems using these libraries are not able to gather semantics of gestures. They also can not predict user intention before the user hand reaches the object.

# **3 Organization of the Related Work**

## **3.1 Organizing through a Real Scenario**

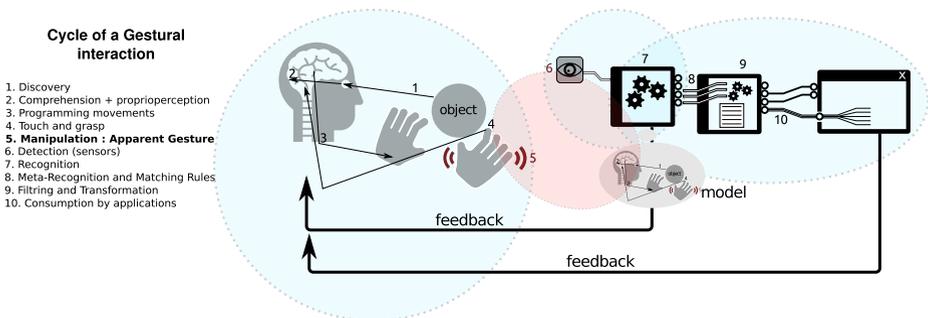
The multidisciplinary nature of the related work makes studying the subject of gesture interaction difficult. Many researchers tried to take the problem from

their point of view. Even with such diverse work, we can start to find what are the common things between them all and then glue parts together. All the research previously referred is around the human and objects. We think that in a real scenario, in which the human is manipulating objects, would put all the reseach pieces in place. Our practical case of study out of this work is with a human manipulating 2D and/or 3D stereoscopic objects which are on or above a table space as in figure 1 and the detailed interaction figure 2.

In this context, the user detects objects through vision with a possible 3D optical illusion (stereoscopy), reaches the object inside his interaction space, grasps the object depending on its form and the intended manipulation, then manipulates it through gestures. In the same time the operator uses 2D and 3D gestures, and this is why a single recognizer for the two cases is needed.



**Fig. 1.** The manipulation of the gestures on and above the table, using 2D and 3D virtual objects and captured via 3D sensors



**Fig. 2.** The cycle of a gestural interaction, including the human, the hand, the object, the sensors, the handling daemons, and applications

### 3.2 Norman's Theory of Human Action as an Organizing Method

From a HCI point of view, and as the human is the generator of gestures, we propose the use of Norman's theory of Human Action Cycle [25] to position the areas of research together on the cycle. Norman describes seven stages-of-action in his model, during an experiment like ours, all these stages are fulfilled as it is included in his description of an interaction. The seven stages can be divided in 3 parts: Goal formation, Execution, and Evaluation. The perception can be in the Goal formation part. The Execution having translation of goals into a set of tasks can be attached to gestemes, the sequencing of tasks into action sequence can be attached to kinemes and reaching, the execution can be through the grasping and manipulation. The evaluation part may include the feedback in a non-exhaustive way, and which is not in the scope of this paper.

### 3.3 The Urge to Link with the HCI Discipline

The goal of connecting areas together means understanding the problem deeper. This understanding can be transformed later into a model able to receive a partial amount of the hand information and then can extrapolate the missing data. We believe that a model based on the shared knowledge from different fields can satisfy the naturality question [12].

### 3.4 The Naturality from EMG and fMRI

The electromyogram (EMG) is the electrical signal detectable by electrodes on the skin of the muscle. And the functional magnetic resonance imaging (fMRI) is a neuroimaging technique that measures brain activity from the changes in blood flow coupled with neuronal activation. The brain activation during an operation allows the detection of what makes an operation stressful. Ehrsson et al. [9] studied forces and brain activation during a power and precision grip manipulation. The results show that even if the power grip forces are higher, the left-sided brain activity is the principal activation and it is low, while in precision grip with small forces generated, both sides and more regions are activated. We think that the more an operation is natural, the less zones are activated in the brain and the less stress it generates.

Other studies confirm the gesture and speech being processed by the same neural system, should they be manipulated and detected using similar methods?

### 3.5 Comparison with a Legacy Device: The Mouse

Gesture technologies have always tried to compare to the mouse. Mainly claiming that they are able to ditch the mouse from its current dominant position in computing. These claims have been demonstrated to be wrong. The mouse is currently more efficient for the usual tasks [3]. The comparison was trying to beat the mouse.

The specificity of the mouse is that manipulating the mouse don't differ a lot from the hand rest position. While gesture systems tend to trigger fatigue and discomfort due to the gorilla arm effect. The rest position of the hand means less activation in the brain and less stress.

### **3.6 Enaction and Vicarious Learning**

There are two ways of organizing knowledge in the brain, either enactive knowledge through action and motor skills like manipulating objects. Or learning by observation. Do we manipulate objects gestures by doing ourselves or by watching others do?

## **4 Prediction of Gestures**

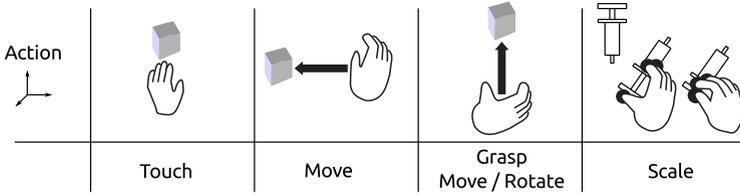
### **4.1 Affordance of Object Grasping**

In a grasp, we can gather rich information in relation to the object properties, the setting, the relationship, the goal, and the anatomy of the user [37]. According to the previous studies on grasping, the way we grasp an object may be predicted according to the opposition plans, and in a virtual environment with virtual objects, we can simplify the model even more by ignoring a part of the grasp properties like the relationship and the goal.

Opposition plans [16,17] gives us hints about where fingers would be positioned for an object. In a virtual environment, we put the focus on pushing the user into a pad opposition as there is no object tactile feedback. This affordance of object grasping prepares the way into manipulation and thus gesture manipulation affordance.

### **4.2 Affordance of Gesture Manipulation**

Manipulating an object in a virtual environment means moving the hands and stroking in the 3D space. The affordance of a manipulative gesture for an object is related to the affordance of grasping on the same object. The prediction of gestures is not always possible. But for many cases, the hand posture before reaching the object is relevant for what the user is willing to do. The form of the object, the hand posture before grasp and the position from where the hand through the limb reaches the object is a telling factor of the user gesture intention. These three, in some cases, remove many alternative possibilities of what is remaining to recognize as in figure 3. A user reaching a pawn object in a chess game from the upper side means he is willing to move it to another place. The scaling of the object is impossible, and the form of the pawn which has a base eliminates the possibility of rotation.



**Fig. 3.** Cases of gestures prediction

## 5 Simplifying of Gesture Manipulation Recognition

### 5.1 Role of the Hand and the Object

In our manipulation study, and in order to simplify the process, we suppose that the hand is the human tool used to act on a object whose existence is required. The notion of naturality of the interaction can not be processed without an object receiving the hand touch, grasp and movements.

### 5.2 Definition of Multitouch Gestures in 4 Classes

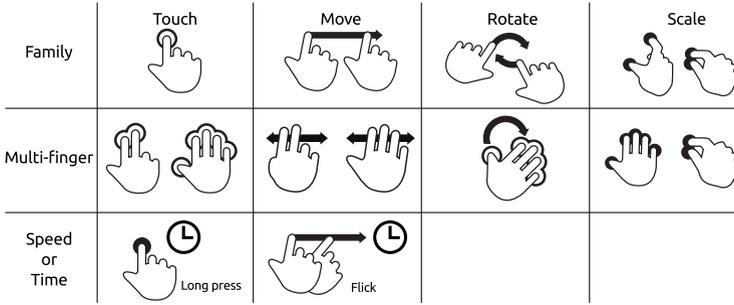
Multi-touch gestures are generally expressed in a wide form of possibilities, each one has a separate name which add a lot of complexity in definition and in use [41]. A burst nomination of multitouch gestures does not facilitate neither their user nor building a recognizer engine for them. With the hypothesis of having an object as a direct receiver of the gesture, we can organize the multi-touch gestures as shown by the figure 4 into four principal families<sup>1</sup>. Then, in order to detect sub-categories, we compute properties like the number of fingers, speed and time. The factorization of gestures into a small set of categories is previously cited by a work of Reisman et al. [30] as Rotate-Scale-Translate interaction.

### 5.3 Definition of Spatial Gestures in 4 Classes

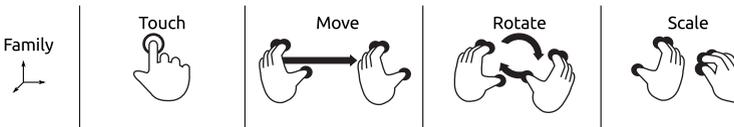
We have questioned spatial gestures classes after handling multi-touch ones. We wanted to know whether it is possible to do the same classification into basic families as in figure 5. The work by Bullock et al. [7] made a specific classification for human manipulation, but using only rotation and translation. In his work, squeezing the syringe is considered as a translation task. We take the object into consideration, and if it is deformable, then a new category is required and in this case is the scaling gesture.

Meanwhile, not all gestures that exist can be expressed using one of the classes if the hypothesis of the receiver object is not satisfied. Non natural gestures that cannot be guessed unless being taught in advance like the Cyclostar by Malacria et al. [18] for dragging and zooming cannot be expressed by our system.

<sup>1</sup> Using multi-touch figures from <http://www.lukew.com/ff/entry.asp?1073>



**Fig. 4.** Families of multi-touch gestures and specific gestures depending on number of fingers or the speed



**Fig. 5.** Families of spatial gestures

## 5.4 Detecting Specific Gestures and the Points of View

Four classes of gesture is not enough to detect all gestures. We need more detailed characteristics allowing fine differentiation. From the characteristics we can list the number of fingers, the speed of the gesture, the delay in operation, the size of the interaction zone, the central point.

How a gesture recognition is handled differ from one system to another. Many systems wait for the gesture to start, record its stroke until the gesture finishes, then starts the recognition. The recognition results are only available at the end. Other more advanced systems, as the ones implemented in smartphones, don't wait until a gesture finishes, they start recognition when a fixed threshold is reached and each chunk of small movement beyond the threshold is decided from the four basic classes. Sometimes many decisions are fired at the same time, letting the final decision to meta recognition tools. In the case of Linux, recognition of long strokes as standalone gesture needs a meta system capable of matching predefined rules with the chunks feed.

## 6 Conclusion and Future Work

In this paper, we have explored most of the areas of research around gestures, in an attempt to extend the view of the problem to a wider public. We have shown that the recognition of gestures needs first a recognition of objects properties, human grasping, understanding of how we see and decide. The context and scenario of a gesture manipulation experience and the Norman's theory of human

action helped us organize these blocks of research and raise the question why some legacy devices still beat new sensors and input devices. We have presented why the prediction of user gestures from hand posture during the grasping can improve current gesture recognizers. We have presented a taxonomy for classifying gestures either in 2D or in 3D in a few main classes and then detect sub-categories using a set of properties. Future work will target a deeper study to strengthen the links between the presented research areas and the development of a real prototype to experiment the theoretical notions presented.

## References

1. Aglioti, S., DeSouza, J.F., Goodale, M.A.: Size-contrast illusions deceive the eye but not the hand. *Current Biology*: CB 5(6), 679–685 (1995)
2. Beaudouin-Lafon, M.: Instrumental interaction: An interaction model for designing post-wimp user interfaces. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 446–453. ACM (2000)
3. Bérard, F., Ip, J., Benovoy, M., El-Shimy, D.: Did Minority Report get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task. *IFIP* (2009)
4. Bertranne, D.: Praxies idéomotrices corporelles: Création d'un test d'imitation de postures asymboliques (2007)
5. Bruno, N., Bernardis, P.: Dissociating perception and action in Kanizsa's compression illusion. *Psychonomic Bulletin & Review* 9(4), 723–730 (2002)
6. Bullock, I., Ma, R., Dollar, A.: A Hand-Centric Classification of Human and Robot Dexterous Manipulation. [ieeexplore.ieee.org](http://ieeexplore.ieee.org), section III, 1–16 (2012)
7. Bullock, I.M., Dollar, A.M.: Classifying human manipulation behavior. In: *2011 IEEE International Conference on Rehabilitation Robotics (ICORR)*, pp. 1–6. IEEE (2011)
8. Cutkosky, M.: On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on Robotics and Automation* 5(3), 269–279 (1989)
9. Ehrsson, H.H., Fagergren, A., Jonsson, T., Westling, G., Johansson, R.S., Forssberg, H.: Cortical activity in precision-versus power-grip tasks: An fmri study. *Journal of Neurophysiology* 83(1), 528–536 (2000)
10. Feix, T., Pawlik, R., Schmiedmayer, H.-B., Romero, J., Kragic, D.: A comprehensive grasp taxonomy. In: *Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation*, pp. 2–3 (2009)
11. Frohlich, B., Tramberend, H., Beers, A., Agrawala, M., Baraff, D.: Physically-based manipulation on the responsive workbench. In: *Proceedings of the IEEE Virtual Reality*, pp. 5–11. IEEE (2000)
12. Gamberini, L., Spagnoli, A., Prontu, L., Furlan, S., Martino, F., Solaz, B.R., Alcañiz, M., Lozano, J.A.: How natural is a natural interface? An evaluation procedure based on action breakdowns. *Personal and Ubiquitous Computing* (October 2011)
13. Gustafson, S., Bierwirth, D., Baudisch, P.: Imaginary interfaces: Spatial interaction with empty hands and without visual feedback. In: *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology*, pp. 3–12. ACM (2010)
14. Harrison, C., Schwarz, J., Hudson, S.E.: Tapsense: Enhancing finger interaction on touch surfaces. In: *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, pp. 627–636. ACM (2011)

15. Hilliges, O., Izadi, S., Wilson, A., Hodges, S., Garcia-Mendoza, A., B.: Interactions in the air: Adding further depth to interactive tabletops. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, pp. 139–148. ACM (2009)
16. Iberall, T., Bingham, G., Arbib, M.: Opposition space as a structuring concept for the analysis of skilled hand movements. *Experimental Brain Research Series* (1986)
17. MacKenzie, C.L.C., Iberall, T.: *The grasping hand*. Elsevier (1994)
18. Malacria, S., Lecolinet, E., Guiard, Y.: Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: The cyclostar approach. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2010, pp. 2615–2624. ACM, New York (2010)
19. Marzke, M.W., Wullstein, K.L.: Chimpanzee and human grips: A new classification with a focus on evolutionary morphology. *International Journal of Primatology* 17(1), 117–139 (1996)
20. McNeill, D.: *Gesture and thought*. University of Chicago Press (2008)
21. Mine, M.R., Brooks, J. F.P., Sequin, C.H.: Moving objects in space: Exploiting proprioception in virtual-environment interaction. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1997, pp. 19–26. ACM Press/Addison-Wesley Publishing Co., New York (1997)
22. Nacenta, M.A., Kamber, Y., Qiang, Y., Kristensson, P.O.: Memorability of pre-designed and user-defined gesture sets. In: CHI, pp. 1099–1108 (2013)
23. Napier, J., Tuttle, R.: *Hands*. Natural science. Princeton University Press (1993)
24. Napier, J.J.: The prehensile movements of the human hand. *Surger* 38(4), 902–913 (1956)
25. Norman, D.: *The design of everyday things* (2002)
26. Norman, D.: Natural user interfaces are not natural. *Interactions*, 6–10 (2010)
27. Norman, D.A., Draper, S.W.: *User Centered System Design; New Perspectives on Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale (1986)
28. Nowak, D., Hermsdörfer, J.: *Sensorimotor Control of Grasping: Physiology and Pathophysiology*. Cambridge University Press (2009)
29. Paulignan, Y., MacKenzie, C., Marteniuk, R., Jeannerod, M.: Selective perturbation of visual input during prehension movements. *Experimental Brain Research* 83(3), 502–512 (1991)
30. Reisman, J.L., Davidson, P.L., Han, J.Y.: A screen-space formulation for 2D and 3D direct manipulation. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, UIST 2009, p. 69 (2009)
31. Roberts, A.I., Vick, S.-J., Roberts, S.G.B., Menzel, C.R.: Chimpanzees modify intentional gestures to coordinate a search for hidden food. *Nature Communications* 5 (2014)
32. Roffman, I., Savage-Rumbaugh, S., Rubert-Pugh, E., Ronen, A., Nevo, E.: Stone tool production and utilization by bonobo-chimpanzees (*pan paniscus*). *Proceedings of the National Academy of Sciences* 109(36), 14500–14503 (2012)
33. Sève-Ferrieu, N.: *Neuropsychologie corporelle, visuelle et gestuelle: Du trouble à la rééducation*. Elsevier Masson (2005)
34. Valkov, D.: *Interscopic Multi-Touch Environments*. dfki.de, 339–342 (2010)
35. Victor, B.: *A Brief Rant on the Future of Interaction Design* (2011), <http://worrydream.com/ABriefRantOnTheFutureOfInteractionDesign/>
36. Vinayavekhin, P.: *Dexterous manipulation planning from human demonstration*. PhD thesis, University of Tokyo (2009)

37. Wimmer, R.: Grasp sensing for human-computer interaction. In: Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction, pp. 221–228. ACM (2011)
38. Wing, A., Haggard, P.: Hand and brain: The neurophysiology and psychology of hand movements
39. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2009, pp. 1083–1092. ACM, New York (2009)
40. Xu, J., Gannon, P.J., Emmorey, K., Smith, J.F., Braun, A.R.: Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences* 106(49), 20664–20669 (2009)
41. Yee, W.: Potential limitations of multi-touch gesture vocabulary: Differentiation, adoption, fatigue. In: Jacko, J.A. (ed.) *HCI International 2009, Part II*. LNCS, vol. 5611, pp. 291–300. Springer, Heidelberg (2009)