Natural User Interfaces: Why We Need Better Model-Worlds, Not Better Gestures.

Hans-Christian Jetter

University of Konstanz

hans-christian.jetter@uni-konstanz.de

Jens Gerken

University of Konstanz

jens.gerken@uni-konstanz.de

Harald Reiterer

University of Konstanz

harald.reiterer@uni-konstanz.de

Abstract

We introduce our view of the relation between *symbolic gestures* and *manipulations* in multi-touch Natural User Interfaces (NUI). We identify *manipulations* not gestures as the key to truly natural interfaces. Therefore we suggest that future NUI research should be more focused on designing visual workspaces and model-world interfaces that are especially appropriate for multi-touch manipulations.

Keywords

Gestures, Manipulations, Direct Manipulation, Conversation Metaphor, Model-World Metaphor.

Introduction

Natural User Interfaces (NUI) promise to introduce more natural ways of interacting with computers into our professional and private life. To achieve this, many NUI designers and researchers are focused on creating and evaluating natural gesture sets for multi-touch interaction (e.g. [7]) and improving the gestures' visual feedback and learnability. Without any doubt, these efforts are highly relevant and suit the urgent market need to quickly introduce gestures in our existing operating systems. We believe however, that this focus

on gestures has to be carefully balanced with a candid reflection about the cognitive aspects and realistic prospects of NUIs and touch gestures. This includes holistic considerations of NUI input *and* output that also take into account the many documented shortcomings of today's desktop metaphor, Graphical User Interface (GUI) and WIMP (Windows Icons Menu Pointer) inter-action style. Focusing on gestural input without funda-mental changes to the structure and visualization of content & functionality will not bring us closer to the promise of truly natural computing. For innovative NUIs, we must therefore create new visual model-worlds based on appropriate visual metaphors, visual formalisms, and coherent conceptual models in which we can act naturally using manipulations. Today's mazelike WIMP interface with occluding windows, walled applications, and a restrictive file system must not stay the dominant model, if NUIs shall lead us towards a new era of natural interaction. To further clarify our position, we introduce three theses about the design of NUIs in the following that reflect our experiences and will explain our conclusion.

Manipulations are not gestures.

We believe in a fundamental dichotomy of multi-touch gestures on interactive surfaces. This dichotomy differentiates between two classes of multi-touch interactions: *symbolic gestures* and *manipulations.*

For us, *symbolic gestures* are close to the keyboard shortcuts of WIMP systems. They are not continuous but are executed by the user at a certain point of time to trigger an automated system procedure. There is no user control or feedback after triggering. In Windows 7 for example, flicking your finger left or right will execute a jump forward or backward in browser history. In [7] writing a check symbol ('✓') or a cross-out symbol ('✗') with the finger is a gesture for "accept" or "reject". Future systems will introduce more of such gesture-based shortcuts that trigger actions like "maximize window" or "change pen size". Our notion of symbolic gestures has been inspired by discussions among NUI practitioners on the Web. For example, designer Ron George also differentiates between "gestures" and "manipulations". For him gestures are indirect: "they do not affect the system directly according to your action. Your action is symbolic in some way that issues a command, statement, or state." [2].

The opposite class of multi-touch interactions is *manipulations*. Unlike symbolic gestures, manipulations are continuous between manipulation initiation (e.g. user fingers down) and completion (e.g. user fingers up). During this time span, user actions lead to smooth continuous changes of the system state with immediate output. Typical examples of manipulations are the dragging, resizing, and rotating of images. Further typical manipulations can be found in Geographical Information Systems (e.g. Microsoft Virtual Earth), in the iPhone Web browser Safari, or in our ZOIL widgets for Zoomable User Interfaces (ZUI) for the Microsoft Surface SDK [5]: users can smoothly control the zoom level by pinching or spreading a zoomable canvas and pan by sliding on it with their fingers. Our understanding of manipulations is based on the traditional principles of *direct manipulation* as introduced by Shneiderman in context of the GUI in 1982 [6]: "Continuous representation of the object of interest" and "rapid incremental reversible operations whose impact on the object of interest is immediately visible". Shneiderman's third principle "physical actions

or labeled button presses instead of complex syntax" is still relevant regarding the complex "syntax" that some symbolic gestures have (e.g. whole sequences of touching and sliding with multiple fingers).

Interestingly, our above examples of manipulations would be regarded as gestures by some authors in HCI, while others would not regard them as gestures at all. Particularly a linguistic or anthropological view would not consider manipulations as gestures, since they are not symbols and part of human communication but intentional physical activities to change the surrounding world. In this sense, gestures would always mean "to talk about (intended) things", while manipulations mean "to act in the world to make (intended) changes".

Manipulations are natural. Gestures are not.
Hutchins et al.'s *conversation metaphor* and *model-world metaphor* from their cognitive account of direct manipulation interfaces from 1985 [3] are helpful in further understanding NUIs. The two metaphors describe opposite mental models that creators of user interfaces can have about the nature of human-computer interaction: "In a system built on the conversation metaphor, the interface is a language medium in which the user and system have a conversation about an assumed, but not explicitly represented world. In this case, the interface is an implied intermediary between the user and the world about which things are said. In a system built on the model-world metaphor, the interface is itself a world where the user can act, and which changes state in response to user actions. The world of interest is explicitly represented and there is no intermediary between user and world. Appropriate use of the model-world metaphor can create the sensation in the user of

acting upon the objects of the task domain themselves." [3]

We believe that symbolic gestures are the reincarnation of the conversation metaphor in NUIs. Symbolic gestures are indirect and resemble learning an artificial sign language to converse with a system about triggering actions. Symbolic gestures alienate the NUI from the key achievement of the GUI: direct manipulation. Ironically the integration of more multi-touch gestures into our WIMP operating systems and applications will lead to a generation of pseudo-natural user interfaces which are in many respects closer to the command line interface (CLI) than to intuitive GUIs or NUIs. In his blog, software consultant Joshua Blake therefore criticizes such "system gestures" for being "not really NUI worthy. They are actually a step backward into the era of rote learning of CLI commands, except with touch" [1].

Not surprisingly, we therefore consider non-symbolic direct manipulation in a model-world as the key to truly natural interfaces. Properly designed manipulations in a logical and consistent model-world can minimize the *semantic distance* between user goals and the available user actions. They thereby create a feeling of directness or direct engagement that "results from the commitment of fewer cognitive resources" [3]. Their logical consistency and fault tolerance invites to explore more functionality and increases learnability. Other HCI concepts, such as Jacob et al.'s *reality-based interaction* [4] can thereby be used to inspire the design of such a model-world and its allowable manipulations, e.g. they can suggest drawing from the users' pre-existing body or environment awareness & skills or the users' understanding of naïve physics.

We need better model-worlds.

We are convinced that a good NUI invites the user to immerse into a simple visual model-world representing the application-domain. However, there are not many convincing model-worlds available for NUIs yet: *WIMP approaches* remain too close to today's GUI and only complement it with a set of additional manipulations and symbolic gestures. These insular improvements can hardly compensate for the mazelike WIMP interface and its many shortcomings, so that such NUIs can never become truly natural. *3d virtual workspaces* use sophisticated graphics and physics engines to create a stunning visual appearance and the perfect illusion of physical behavior (e.g. rotation, inertia, collision…). However, unless used for 3d visualization purposes, such interfaces often fail to efficiently provide any functionality that goes beyond the navigation and movement of objects in space. Furthermore, the third dimension adds considerable complexity to view management and navigation on a flat interactive surface. The *scatter view*, a two-dimensional plane in which objects can be freely dragged, rotated, and resized is probably the most popular model-world in today's NUIs. However, scatter views are inappropriate for most applications as they are limited to the screen size and cannot handle multiple items without great overlaps or shrinking the items' size to few pixels.

We suggest that future NUI research should be more focused on the design of better model-worlds for realistic use-cases. These model-worlds must support large numbers of items, integrate live Web pages and documents, and support visual search & filtering. An initial step into this direction is our ZOIL (Zoomable Object-Oriented Information Landscape) paradigm and software framework which introduces a zoomable canvas and semantic zooming to the Surface SDK [5]. We will be happy to demonstrate ZOIL and its use-cases to trigger a discussion with the other workshop participants about future research directions and real-world requirements for NUI model-worlds.

Acknowledgements

References

[1]   Blake, J. Deconstructing the NUI. http://nui.joshland.org/

[2]   George, R. OCGM (pronounced Occam['s Razor]) is the replacement for WIMP. http://blog.rongeorge.com/

[3]   Hutchins, E.L., Hollan, J.D., and Norman, D.A. 1985. Direct manipulation interfaces. *Hum.-Comput. Interact. 1, 4* (1985), 311-338.

[4]   Jacob, R.J., Girouard, A., Hirshfield, L.M., Horn, M.S., Shaer, O., Solovey, E.T., and Zigelbaum, J. Reality-based interaction: a framework for post-WIMP interfaces. In *Proc. CHI 2008*, ACM Press (2008), 201-210.

[5]   Jetter, H.-C., König, W., and Reiterer, H. Understanding and Designing Surface Computing with ZOIL and Squidy. Multitouch and Surface Computing (CHI 2009 Workshop), Apr 2009.

[6]   Shneiderman, B. The future of interactive systems and the emergence of direct manipulation. *Behavior and Information Technology, 1* (1982), 237-256.

[7]   Wobbrock, J.O., Morris, M.R., Wilson, A.D. User-defined gestures for surface computing. In *Proc. CHI 2009*, ACM Press (2009), 1083-1092.