

*Co-Optjects:
Instant User
Interfaces Through
Everyday Objects*

Master's Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University



*by
Christian Corsten*

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. James D. Hollan, Ph.D.
Registration date: Jul 12th, 2011
Submission date: Jan 30th, 2012

I hereby declare that I have created this work completely on my own and used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Aachen, January 2012
Christian Corsten

Contents

Abstract	xix
Überblick	xxi
Acknowledgements	xxiii
Conventions	xxv
1 Introduction	1
1.1 Appropriation for Digital Interfaces	3
1.2 Towards Instant User Interfaces	6
1.2.1 Benefits of Instant User Interfaces . . .	9
1.3 About Tangible, Organic, and Ubiquitous User Interfaces	11
1.3.1 Tangible User Interfaces (TUIs)	11
1.3.2 Organic User Interfaces (OUIs)	13
1.3.3 Ubiquitous Computing (UbiComp) . .	14
1.4 Research Questions	15
1.5 Outline	16

2	Related Work	19
2.1	Design Space	21
2.1.1	Kind of Manipulation	21
	Direct Manipulation	21
	Indirect Manipulation	22
2.1.2	Types of Gesture	22
	Stationary Gestures	23
	Non-stationary Gestures	24
2.1.3	Exploitation of Physical Affordances .	24
2.1.4	Co-Objects in the Design Space . . .	25
2.2	Related Research Projects	25
2.2.1	Mouseless: <i>Mistry and Maes, 2011</i> . . .	26
2.2.2	Skinput: <i>Harrison et al., 2011</i>	28
2.2.3	WUW – Wear Ur World: <i>Mistry et al., 2009</i>	30
2.2.4	Cloth Displays: <i>Lepinski and Vertegaal, 2011</i>	32
2.2.5	OnObject: <i>Chung et al., 2010</i>	34
2.2.6	DisplayObjects: <i>Akaoka et al., 2010</i> . .	36
2.2.7	SketchSpace: <i>Holman and Benko, 2011</i> .	39
2.2.8	iCon: <i>Cheng et al., 2010</i>	42
2.3	A Futuristic Perspective on Everyday Object Interaction	45
2.3.1	Invoked Computing: <i>Zerroug et al., 2011</i>	45

2.4	Related Work at a Glance	47
3	Diary Study: A Primer To Everyday Object Interaction	51
3.1	Procedure	52
3.2	Participants	53
3.3	Results	53
4	Wizard of Oz Study: Interaction Patterns for Instant User Interfaces	57
4.1	Setup	58
4.2	Procedure	59
4.2.1	Objects	62
4.2.2	Scenarios	64
	Block A: Remote Target Systems . . .	64
	Block B: UI Widgets in GUIs	64
4.3	Participants	65
4.4	Evaluation	66
4.4.1	Method of Analysis	66
4.4.2	Scenario Analysis	67
4.4.3	Scenarios A: Observations	67
	A1: Holding a Presentation	67
	A2: Watching Television	70
	A3: Playing a Video Game	72

	A4: <i>Operating the Light</i>	76
	A5: <i>Interacting with a Virtual Shop Window</i>	78
4.4.4	Scenarios B: Observations	80
	B1: Buttons	80
	B2: <i>Radio Buttons</i>	82
	B3: <i>Check Boxes</i>	83
	B4: <i>Rotary Buttons</i>	85
	B5: <i>Sliders</i>	86
	B6: <i>Scroll Bars</i>	89
	B7: <i>Text Boxes</i>	91
4.4.5	Qualitative User Feedback	92
4.5	Lessons Learned	97
5	Testing Everyday Object Interaction with an Interactive Prototype	103
5.1	Goals	104
5.2	Prototype Design	106
5.2.1	Breathing Life into Passive Objects . .	106
	Touch Detection on Arbitrary Surfaces	107
	Recognition of Stationary and Non-Stationary Movement	111
5.2.2	Linking Objects to Gestures and Actions	112
	A: Kinect Controller	113

B: Fiducial Controller	115
C: Event Processor	117
5.2.3 Applications	118
Television Application	118
QuickTime Player Application	120
5.3 Qualitative User Study	120
5.3.1 Setup	121
5.3.2 Participants	123
5.3.3 Tasks	124
5.3.4 Procedure	126
5.4 Prototype Limitations	129
5.5 Findings	129
5.5.1 Users' Performance on Tasks	130
Task Performance for the Movie Player Scenario	131
Task Performance for the Television Scenario	136
5.5.2 Users' Interpretations	143
5.5.3 Mixing Everyday Object Usage with Appropriation	146
Deactivation Patterns	152
5.5.4 Overall System Evaluation	155
5.6 Lessons Learned	157

6	Summary	161
6.1	Summary and Contributions	161
6.2	Future Work	163
7	Instant User Interface Guidelines	167
7.1	Recommendations and Guidelines	167
A	Appendix for the Wizard of Oz Study	173
A.1	Scenario Sketches	173
A.2	Feedback Form	173
B	Appendix for the Interactive Prototype Study	189
B.1	Feedback Form	189
	Bibliography	193
	Index	197

List of Figures

1.1	A cutlery knife contrasted with a letter opener	1
1.2	A parcel with a drawable leash	4
1.3	Norman's four Stages of Execution	8
1.4	Interaction model of a TUI	12
1.5	Coherence continuum for TUIs	13
2.1	A design space for related work	20
2.2	Physical setup for Mouseless	27
2.3	The Skinput prototype armband	29
2.4	Inout accuracy achieved by Skinput	30
2.5	The WUW hardware prototype	31
2.6	The Cloth Displays prototype	33
2.7	End user programming with OnObject	35
2.8	Hardware setup for DisplayObjects	37
2.9	Adding widgets to prototypes	38
2.10	DisplayObjects credit card interface	39

2.11	The SketchSpace hardware setup	41
2.12	A tilt-enabled computer mouse	42
2.13	The iCon end user programming interface . .	44
2.14	The Invoked Computing hardware prototype	46
3.1	Sample picture for the diary study	53
4.1	Object variants for a drinking vessel	58
4.2	Study setup for the Wizard of Oz study . . .	60
4.3	Incremental visual task feedback for the TV scenario	61
4.4	Mobile phone and remote control	69
4.5	Precondition for scenario A3	73
4.6	Applied Physical Instantiation for scenario A3	76
4.7	Methods for focus switching	88
4.8	Preferred quantity of objects for solving a scenario	95
4.9	Match between selectable objects and objects available in reality	96
4.10	Perceived ease of finding objects and apply- ing suitable gestures on them	97
4.11	Everyday object manipulation (1/2)	100
4.12	Everyday object manipulation (1/2)	101
5.1	The prototype aluminum mounting	108
5.2	Kinect teardown	108

5.3	Colored depth map	109
5.4	Touch detection through depth calculation	110
5.5	A reacTIVision fiducial	112
5.6	Object activation via pointing	113
5.7	Prototype Architecture	114
5.8	Principal investigator's control panel	115
5.9	Touch detection for a stapler	116
5.10	TV channel list and menu	119
5.11	QuickTime Player GUI	121
5.12	Study setup	122
5.13	Everyday objects available for the user study	123
5.14	Predefined gestures for appropriated objects	128
5.15	Glue stick in action (timeline control)	134
5.16	Bottle in action (volume control)	138
5.17	Tissues in action (channel control)	140
5.18	Bottle in action (brightness control)	141
5.19	User-rated intuitiveness of predefined gestures at a glance	144
5.20	Example of an interrupt	147
5.21	Histogram for deactivation preference	150
A.1	A1: Holding a Presentation	174
A.2	A2: Watching Television	175

A.3	A3: Playing a Video Game	176
A.4	A4: Operating the Light	177
A.5	A5: Interacting with a Virtual Shop Window	178
A.6	B1: Buttons	179
A.7	B2: Radio Buttons	180
A.8	B3: Check Boxes	181
A.9	B4: Rotary Buttons	182
A.10	B5: Sliders	183
A.11	B6: Scroll Bars	184
A.12	B7: Text Boxes	185
A.13	WOZ Feedback Form, pages 1 & 2	186
A.14	WOZ Feedback Form, pages 3 & 4	187
B.1	Interactive Prototype Feedback Form, pages 1 & 2	190
B.2	Interactive Prototype Feedback Form, page 3	191

List of Tables

- 2.1 Related research projects at a glance (1/2).
 Legend: – not/little, ○ medium, + strong,
 ☛ touch, □ non-stationary 2D movement,
 ■ stationary and non-stationary 2D move-
 ment, ▣ 3D movement, ➡ acceleration, ↔
 deformation 48
- 2.2 Related research projects at a glance (2/2).
 Legend: – not/little, ○ medium, + strong,
 ☛ touch, □ non-stationary 2D movement,
 ■ stationary and non-stationary 2D move-
 ment, ▣ 3D movement, ➡ acceleration, ↔
 deformation 49
- 3.1 Object occurrences identified for the diary
 study. This table shows the number of ob-
 ject category occurrences in the given envi-
 ronment (work vs. leisure). Whenever an ob-
 ject belonging to a category was observed at
 least once for a participant, the correspond-
 ing leisure/work counter was incremented
 ($N = 19$). The table is ranked by the over-
 all occurrences and shows only items that
 occurred overall at least five times based on
 qualitative counting. 55

-
- 4.1 Available object types per scenario. Objects for scenarios A1–A4 were based on the most frequently observed items available in leisure environments according to the diary study; objects for scenario A5 represented wearable items and were a subset of objects used in A1–A4. Objects for scenarios B1–B7 were based on items at hand in work environments. The remote control was not available in scenarios A1 and A2. PC mouse and PC keyboard were only allowed to be used as general multi-button devices. 63
- 4.2 Object properties preferred for appropriation. Users preferred small and lightweight objects equipped with buttons or controls. As regards rigidity/flexibility, no preference could be identified. 94
- 5.1 **Tasks and solutions for the TV scenario.** The table presents the demanded tasks (T) alongside with a description concerning how an object had to be manipulated to achieve the goal. Alternating interrupts (I) were added to simulate normal object usage. Letters in brackets indicate alternatives. 125
- 5.2 **Tasks and solutions for the QuickTime Player scenario.** The table presents the demanded tasks (T) alongside with a description concerning how an object had to be manipulated to achieve the goal. Alternating interrupts (I) were added to simulate normal object usage. Letters in brackets indicate alternatives. 126

-
- 5.3 Task success for both scenarios. In general, users were more successful at controlling the movie player than the TV. As regards the movie player, only entering/exiting fullscreen was problematic for three users. For the TV scenario, altering the brightness was most challenging, whereas all users ($N = 16$) managed volume control. . 132
- 5.4 **Deactivation patterns observed by quantity.** For both deactivation methods the distribution was nearly equal for the two scenarios (± 1 user, $N = 16$). As regards pointing, most testers immediately deactivated an object before using another one. Patterns for hiding showed no preference. 153

Abstract

Inspired by the process of appropriation of everyday objects (e.g., a knife) as workaround for replacing specially designed tools (e.g., a letter opener), we transferred this paradigm to the field of digital interfaces. We introduce a new interaction metaphor, called *Instant User Interfaces* (IUIs), which enables people to spontaneously use everyday artifacts as temporary replacements in case the dedicated input device is missing.

As initial approach, we conducted a diary photo study for determining which objects are perceived “everyday” and therefore come into consideration for spontaneous appropriation. Next, in the scope of a Wizard of Oz study, we tested hypothetic appropriation of these objects for various digital interfaces encountered daily. For each scenario, we identified object-gesture-action mappings as devised by the testers. We found out that concerning discrete input, users tended to imitate a dedicated controller as well as possible with everyday objects. Concerning continuous input, however, people created a physical instance of the target itself, denoted by *Physical Instantiation* (PI). In a final study, we tested everyday object interaction with an interactive prototype technically based on a combination of depth camera and fiducial tracking for turning passive objects into physical controllers. Besides testing predefined mappings based on the results of the previous study, we examined the interplay of normal object usage with appropriation.

The working prototype showed that the idea of IUIs is promising as appropriation of everyday objects helps managing simple tasks without being dependent on dedicated controllers. Despite interaction flaws that still exist due to technical limitations, we see the foundation laid for this new interaction metaphor. Motivated by these results, we present a list of guidelines based on the experience collected in the user studies that serves as a primer to future IUI research.

Überblick

Inspiziert durch die Zweckentfremdung von Alltagsgegenständen (z.B. ein Messer) um mit ihrer Hilfe ein speziell entwickeltes Werkzeug (z.B. ein Brieföffner) durch ein anderes zu ersetzen, haben wir uns diesem Paradigma auf der Ebene digitaler Benutzerschnittstellen angenommen. Wir stellen dabei die neue Interaktionsmetapher *Instant User Interfaces* (IUIs) vor, die die spontane und temporäre Nutzung von Alltagsgegenständen als Ersatz für fehlende dedizierte Eingabegeräte ermöglicht.

Zunächst haben wir eine auf Fotos basierende Tagebuchstudie durchgeführt um herauszufinden, welche Objekte uns täglich umgeben, um diese dann spontan zur Zweckentfremdung nutzen zu können. Danach haben wir eine Wizard of Oz-Studie durchgeführt, in der wir die hypothetische Zweckentfremdung der zuvor ermittelten Gegenstände in Bezug auf uns täglich begegnende digitale Benutzerschnittstellen getestet haben. In Bezug auf Geräte mit diskreter Eingabefunktion stellten wir fest, dass die Nutzer dazu neigten, das dedizierte Eingabegerät so gut wie möglich mit Alltagsgegenständen zu imitieren. Hinsichtlich kontinuierlicher Eingabefunktionen tendierten die Nutzer dazu, die eigentliche Benutzerschnittstelle mit Hilfe von Objekten nachzubauen. Dieses Phänomen bezeichnen wir als *Physical Instantiation* (PI). In einer abschließenden Studie haben wir die Interaktion mit Alltagsgegenständen anhand eines interaktiven Prototypen getestet, der technisch auf einer Kombination aus Tiefenkamera und Objektverfolgung basiert, um passive Gegenstände in Eingabegeräte zu verwandeln. Neben der Verwendung der zuvor ermittelten Gesten haben wir das Zusammenspiel von eigentlichem Objektnutzen und seiner Zweckentfremdung untersucht.

Der funktionierende Prototyp hat gezeigt, dass das Konzept der IUIs vielversprechend ist, da die Zweckentfremdung von Alltagsgegenständen es ermöglicht, einfache Aufgaben zu bewältigen, ohne auf dedizierte Eingabegeräte angewiesen zu sein. Trotz noch existierender Interaktionsprobleme aufgrund technischer Hindernisse sehen wir den Grundstein für diese neue Interaktionsmetapher gelegt. Motiviert durch diese Ergebnisse präsentieren wir eine Liste von Leitlinien basierend auf den Ergebnissen der Nutzerstudien, die als Einführung für zukünftige Forschung bezüglich der IUIs dienen.

Acknowledgements

First of all, I would like to thank all testers who participated in the user studies conducted for this work. Thank you for your constructive feedback and the time spent.

Secondly, I would like to thank my supervisor, Dipl.-Inform. Dipl.-Wirt. Inform. Max Möllers, for his support and feedback.

Besides, I would like to thank Chatchavan Wacharamanotham, M.Sc., for additional feedback given to the design of my user studies.

Special thanks go to Prof. Hollan and Nadir Weibel, Ph.D., for having given me the opportunity to intern at the *Distributed Cognition and Human-Computer Interaction Laboratory* at UC San Diego in 2010. The collaboration with them influenced my choice for choosing Prof. Hollan as second examiner. Special thanks also go to Prof. Dr. Jan Borchers and Dipl.-Inform. Malte Weiss who enabled me to take part in the internship.

Finally, I would like to thank my parents for their full support throughout my college years.

Thank you,

– Christian.

Conventions

Throughout this thesis we use the following conventions.

Text conventions

Definitions of technical terms or short excursus are set off in colored boxes.

DEFINITION OF A TERM:
This is an empty definition of a term.

Definition of a Term

Scenarios are set off in green boxes.

PRESENTATION OF A SCENARIO:
This is an empty scenario description.

*Presentation of a
Scenario*

Examples are set off in blue boxes.

PRESENTATION OF AN EXAMPLE:
This is an empty example description.

*Presentation of an
Example*

Implementation symbols are written in typewriter font.

`method()`

The whole thesis is written in American English. For reasons of politeness, unidentified third persons are described in female form. The first person is written in plural form.

Chapter 1

Introduction

In everyday life, people often use special tools that help them to accomplish everyday tasks. Take a letter opener, for instance. The dedicated tool helps us to quickly open a letter without ripping it or the envelope into pieces. But, to be honest, who of us owns and actually uses a letter opener? Is this dedicated tool actually necessary? A cutlery knife, for instance, looks quite similar to a letter opener and indeed can be used identically to open a letter. In addition, a knife is a lot more ubiquitous as it is used by many human beings for a daily ritual: eating. Figure 1.1 shows both objects at a glance.

Dedicated tools help us solving everyday tasks.



Figure 1.1: A cutlery knife (top) contrasted with a letter opener (bottom). Both objects share similar physical properties.

Both tools are made of steel and have a shaft with a relatively narrow, flat, and tapering body. Using Google image search for finding similar objects gives us nail files and

A knife can act as a letter opener.

screwdrivers. Indeed, all these tools have certain properties that makes it possible/suggest to use them as letter openers: The shape and the texture of the shaft suggests to be grasped, whereas the flat and sharp metal body suggests slicing or cutting something. These intuitive and suggested actions are what cognitive scientist Donald A. Norman defines as *Physical Affordances*.

Definition:
Physical Affordances

PHYSICAL AFFORDANCES:

"[...] the term affordances refers to the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used [...]. Affordances provide strong clues to the operations of things." – [Norman, 2002]

Thus, having found an object with physical affordances similar to the letter opener, things like a knife, a nail file, or a screw driver lessen the dependency on the dedicated device and make it unnecessary if not superfluous.

Even a nutcracker can be substituted.

Opening walnuts is quite similar to the example above. Where a nutcracker is missing, any other solid tool resembling a hammer or pliers can help opening the nuts. A stapler, for instance, does a very good job in cracking the shells – proven by ourselves.

These two examples show that people are very good and creative in helping themselves for – at least temporarily – substituting specialized tools in case they are unavailable. This leads to the formal definition of *appropriation*.

Definition:
Appropriation

APPROPRIATION:

The process of using a designed tool or object in a different context and for a different purpose than intended, is called appropriation.

Appropriation is also used for illustrating descriptions, making them easier to understand. In his book "Things That Make Us Smart: Defending Human Attributes in the Age of the Machine" [Norman, 1993], Norman illustrates

how people use common objects used in the office, such as pencils and paper clips, to describe a car crash situation. The pencils are used for illustrating the roads, whereas the paper clips symbolize cars. Recalling the definition of the term “appropriation”, the objects are used in a new context (office vs. illustration) for a different purpose (writing/clipping paper vs. representing huge objects).

Objects can be appropriated in order to serve as small instances by representing huge items.

All examples listed above have in common that people appropriate objects for *mechanical* or *analog* interfaces. Ultimately, this observation leads to a specific question: How about digital interfaces? Can we – similarly to analog interfaces – help ourselves by using common objects for controlling the interface in case the dedicated input device is missing? How can the dependence on dedicated tools be loosened in terms of digital interfaces?

Can the idea of appropriation also be transferred to digital interfaces?

1.1 Appropriation for Digital Interfaces

Of course, one could argue that adding controls directly on the device is a solution to become independent of a separate and dedicated input controller. Take a washing machine, for example. The device is equipped with buttons and knobs mounted on the chassis. Thus, the controls are always at hand. Similarly, this argument can be transferred to analog interfaces. Recalling the envelope, an integrated flap could help overcoming the dependence on a letter opener. Figure 1.2 exemplifies this mechanism for a parcel. Pulling the leash opens the cardboard box without the need for a separate tool. Though being handy, such integrated tools have a significant drawback: Integrating them into objects means that *each* instance has to be specially prepared which results in a much more complex production process alongside with additional costs, special material, and a probably larger form factor.

Integrated tools help becoming independent of separate tools.

Revisiting digital interfaces, separate input devices have two major advantages over integrated controllers: First, special controllers allow to control a device at distance. Take a light switch, for instance. Basically, it can be installed anywhere on the wall. If the control was directly



Figure 1.2: A parcel with an integrated, drawable leash. No separate tool is required to open the paperboard container.

Separate tools are both remote and universal.

mounted on the ceiling lamp, we would always have to jump in order to reach the switch – quite impractical. Second, extra controllers are often universal. Take a computer mouse, for instance. It can be connected to *any* computer to control a graphical user interface which is analog to the letter opener that can open *any* envelope.

Dedicated tools are used in everyday life.

Similarly to the computer mouse, we use dedicated input devices in various scenarios in quotidian life. We operate light switches to dim the light, we use game pads to play video games, or we use a remote control to operate the TV. However, when these controllers are broken or missing, we cannot manipulate the interface. The following sample scenario illustrates how appropriation – similar to the analog world – could help in solving this problem.

Scenario:
Watching Television
(1/2)

WATCHING TELEVISION (1/2):

Peter, aged 44, has just arrived home after work. For workout, he decides to watch a sports program on TV while lying on the couch. Peter is looking for the TV remote control in order to wake up the device from standby mode. Yet, having searched for approximately three minutes, he cannot find the remote control.

WATCHING TELEVISION (2/2):

Being on edge, Peter heads to the TV, pushes the power button, and gets back to the couch. The TV turns on and shows the first channel. However, the sports emission is broadcasted on channel 15. Again, Peter gets up and walks to the television set. He is looking for a channel switch on the the device. To his surprise, there is no such button on the device – just the power switch. Any other command, such as switching the channel, has to be executed using the remote control. Being frustrated, Peter resigns. No sports program today. No relaxing.

*

If only he could use *any* object around him to mimic the remote control. Peter picks a pack of pocket tissues out of his pocket. Its shape reminds him of the shape of a remote control. As the pack is soft, it also feels very comfortable for pushing. He thinks: *“This would be an ideal substitute for quickly switching to the next and previous channel. Pushing the upper half would show to the previous channel; pushing the lower could switch to the next channel.”*

Scenario:
Watching Television
(2/2)

Peter’s intention is obvious. He wants to appropriate the tissues for using them as temporary replacement for the remote control that he could not find. Using the tissues as channel switch would help him reaching his goal: watching the sports emission on channel 15. This *spontaneous* and *instant* co-option of digital functionality to a passive object leads towards the idea of *Instant User Interfaces* (IUIs).

A pack of pocket tissues could serve as remote control replacement, since both objects have a similar shape.

1.2 Towards Instant User Interfaces

We define an *Instant User Interface* (IUI) as follows:

Definition:
*Instant User
Interface (IUI)*

INSTANT USER INTERFACE (IUI):

An Instant User Interface is a user interface that enables the user to select an object within reach in order to control a remote technical system. Based on existing knowledge about physical affordances of the object and the features of the target system, the user establishes ad hoc mappings from artifact to system function without requiring explicit prior setup.

IUIs allow for ad hoc mapping between object and target.

The instant usage of an object as physical controller for a remote target system is – consequently – applied appropriation. Recalling the sample scenario, Peter’s vision of using the tissues as remote controller is an IUI. The pack of tissues is the object within reach, whereas the remote technical system is the television set. To him, the tissues suggest pushing (physical affordances); the ad hoc mapping is the desired link between pushing the tissues and switching the channels.

IUIs let the user design the input controller.

This interaction metaphor takes a new approach towards the user: She is in power of choosing the desired object with the affordances that are most comfortable for controlling the task. Interaction designers, on the contrary, look at the real world and bring these metaphors to the digital world. They design dedicated input devices the user must learn to use. In the context of IUIs, however, the user becomes the designer. She can pick an object she is familiar with and therefore does not need to learn how the object is to be used.

IUIs are taken into consideration when the user cannot reach a certain goal due to the missing dedicated tool.

Instant User Interfaces are taken into consideration when the user cannot reach a certain goal that is linked with a technical remote system. Accomplishing the goal actually precedes a sequence of several actions that have to be executed. At that point, when the user cannot execute these actions that are normally performed with a special tool which is now missing, the concept of IUIs comes to the fore. Identifying the stage when this takes place requires

taking a closer look at the *Stages of Execution* described by Norman [2002].

Norman breaks people's manipulation of the environment down into four Stages of Execution, as depicted in Figure 1.3. First, the user defines a goal, such as "switch to channel 15". The formulation of this goal is imprecise as it does not tell what has to be done for achievement (and how). Therefore, the next stage, the intention to act, is more precise and at least reveals *what* is to be done for achieving the goal, such as "push the channel button on the remote control". This intention is still too vague. Elaboration on the intention is done by formulating a concise sequence of actions: (1) get off the couch, (2) take the remote control, (3) point the remote control at the TV, and (4) push the channel-up button. Finally, for achieving the goal, these actions must be executed by the user. Consequentially, she interacts with the the world:

A user's manipulation of the environment can be broken down into four stages.

The specific actions bridge the gap between what we would like to have done (our goals and intentions) and all possible physical actions. – Norman [2002]

Norman also explains that the same goal often can be achieved by different intentions. So, one might alter the intention, which consequently alters the action sequence, although the goal is still the same. Recalling Peter from the sample scenario, he changed his intention from "push the channel button on the remote control" to "push the channel button on the television set" since he could not find the remote control, inhibiting the execution of the original action sequence. This conflict of intended actions and actually possible actions is identified as the *Gulf of Execution*.

A user can change her intention to act without altering the overall goal.

GULF OF EXECUTION:

"The difference between the intentions and the allowable actions is the Gulf of Execution." – [Norman, 2002]

Definition:
Gulf of Execution

Basically, in terms of digital user interfaces, if a dedicated input device is missing, we have a Gulf of Execution, since there are no allowable/executable actions. Peter could not press the channel button because it was missing: Neither

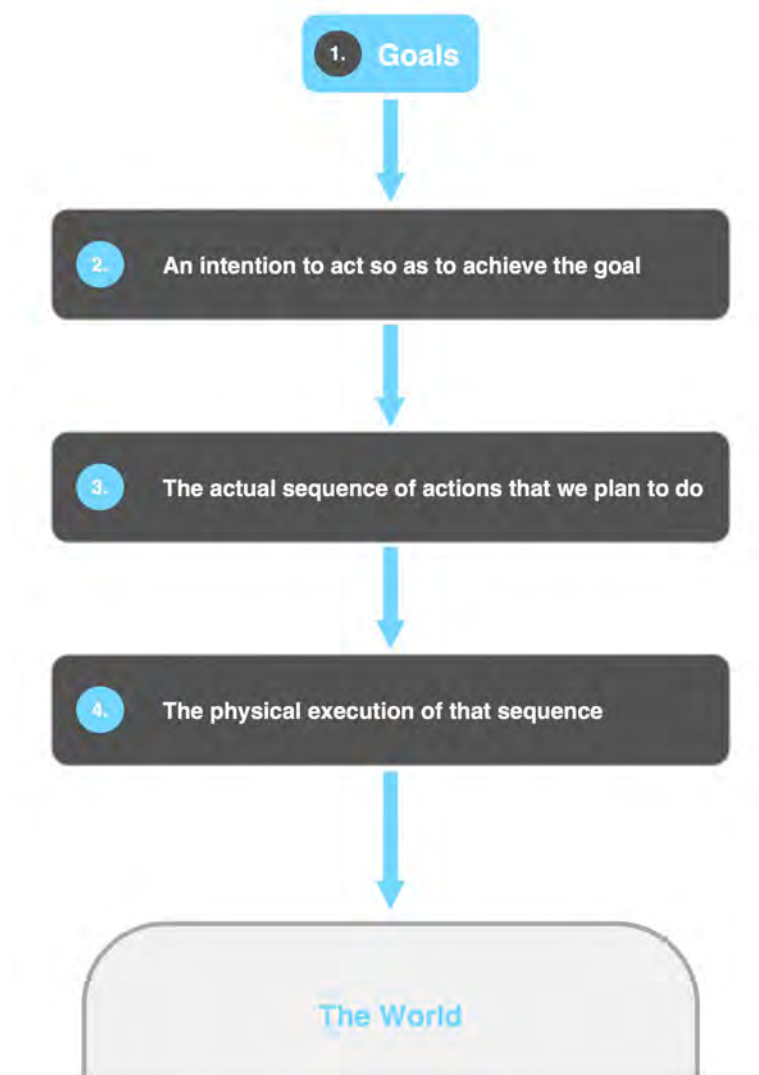


Figure 1.3: Norman's four Stages of Execution. A roughly defined goal is broken down into more and more concise interaction steps.

An unexpectedly missing dedicated input device results in a Gulf of Execution.

the remote control was available nor were there any channel buttons on the TV set. Instant User Interfaces, however, can bridge this Gulf of Execution. They allow the user to improvise by picking any object at hand and by instantly appropriating it as controller. Recalling the sample scenario, the pack of tissues served as temporary controller for switching the channels due to their physical af-

fordances suggesting a remote control-like shape and a push-responsive texture. Hence, appropriating the tissues in the sense of IUIs is a solution to both Gulfs of Execution experienced by Peter. Now, he *can* watch TV. His goal is reached. More specifically, such improvised actions are called *Opportunistic Actions*.

OPPORTUNISTIC ACTIONS:

“Opportunistic actions are those in which the behavior takes advantage of the circumstances.” – [Norman, 2002]

Definition:
*Opportunistic
Actions*

1.2.1 Benefits of Instant User Interfaces

Instant User Interfaces are not targeting at permanently replacing a dedicated input device. As emphasized, spontaneity and instant access are demanded in such case where the dedicated tool is not available or unusable – for whatever reason. As long as the absence of the tool persists, IUIs can serve as *temporary substitutes* for giving the user access to the target system. Basically, IUIs encompass the following benefits:

IUIs serve for temporarily substituting dedicated input devices.

- **Spontaneity.** Being completely unprepared for an incidence where a dedicated input device is missing or broken, IUIs allow last-minute bridging of a Gulf of Execution through opportunistic actions. There is no need to look for an identical replacement which is time-critical in case the tool is needed just in time.

IUIs allow for spontaneously bridging a Gulf of Execution.

Example: Two friends are about to play a video game. Unfortunately, they have just one controller. Since it is late, buying or borrowing a second game pad is impossible. Appropriating an ordinary TV remote control with its physical buttons as controller enables the friends to start playing immediately.

- **Vicinity.** Instead of having to reach for a special controller which may be out of reach, IUIs allow for grabbing any object within reach to take over the functionality of the distant input device.

IUIs are available within reach.

IUIs can be used to map short cuts for GUIs to everyday objects.

Example: Somebody lying on the couch is watching TV and needs to darken the ceiling light. Instead of getting up and reaching for the switch, the person could use the bottle next to her and move it (counter-) clockwise for dimming the light down.

- **Short-Cutting.** Especially in terms of GUIs, IUIs can be used to temporarily cut short sequences of commands to objects for quick access. This is time-efficient and frees the user from remembering awkward keyboard short cut combinations by simultaneously providing eyes-free interaction, mapping one short cut to exactly one object.

Example: A user needs frequent access to an online dictionary while writing a document. Pushing the pocket tissues next to her, the browser window would bring the browser to focus allowing her to search for the word. Tapping again brings the word processor back to focus. Fast focus switching with tactile feedback is guaranteed.

IUIs exploit physical affordances and therefore minimize eye contact with the controller.

- **Eyes-Free Interaction.** Due to object-specific properties, physical affordances can be exploited for using artifacts within reach as tactile controllers suitable to the user's individual needs. Strong tactile cues therefore minimize necessary eye contact with the controller.

Example: A person is watching TV. For controlling the volume she needs to sight the volume buttons on the remote control which is difficult as the room is dark. Turning a bottle for controlling the volume can be done without having eye contact.

IUIs inherit from TUIs, OUIs, and UbiComp.

The examples presented illustrate for which occasions IUIs come into question: (1) when a dedicated device is forgotten, unavailable, or broken, (2) when a dedicated controller is out of reach, and (3) when short cuts are needed. The general benefits of IUIs, i.e., spontaneity and the exploitation of haptic properties for eyes-free control, emerge from three different fields in user interface design; the concept of IUIs creates a link between *Tangible User Interfaces*, *Organic User Interfaces*, and *Ubiquitous Computing*.

1.3 About Tangible, Organic, and Ubiquitous User Interfaces

The IUI concept is inspired by different user interface metaphors. Principally based on Tangible User Interfaces (TUIs), it also reaches towards Ubiquitous Computing (UbiComp) and, to a limited extent, towards Organic User Interfaces (OUIs). We will briefly explain each metaphor and relate IUIs to each field of research.

1.3.1 Tangible User Interfaces (TUIs)

Principally, TUIs target at making digital information tangible [Ishii and Ullmer, 1997]. Where in typical HCI the user interacts with GUIs of virtual nature, TUIs integrate physical objects in the user interface: The world (such as illustrated in Figure 1.3) is the interface. Ishii and Ullmer defined a TUI as user interfaces that “augment the real physical world by coupling digital information to everyday physical objects and environments.”. Fishkin [2004] adds that “they all share the same basic paradigm – a user uses their hands to manipulate some physical object(s) via physical gestures; a computer system detects this, alters its state, and gives feedback accordingly.”. Therefore, TUIs take advantage of physical affordances for giving the user tactile feedback and therefore a richer user experience compared to pure virtual interfaces, “achieving a highlighted legibility and seamlessness of interaction between people and information” [Ishii and Ullmer, 1997]. An example of a TUI for bringing haptic feedback to multi-touch tables is SLAP Widgets [Weiss et al., 2009], as tangibles bridge the gap between virtual and physical controls on tabletops. Unlike pure gestural interaction, physical controls give haptic feedback to the user, resulting in richer user experience. Thus, interaction with physical everyday objects is also superior to pure gesture-based interaction.

TUIs integrate physical objects in the (virtual) user interface.

Figure 1.4 illustrates the interaction model of a TUI proposed by Ullmer and Ishii [2000]. A control, such as a button, is a physical representation of the graspable

For TUIs, virtual controls have physical, graspable counterparts.

virtual counterpart. The model clearly illustrates that physical world and virtual world are directly connected. Manipulation of the physical world affects the virtual one. In traditional GUIs, however, physical and virtual control do not share the same model.

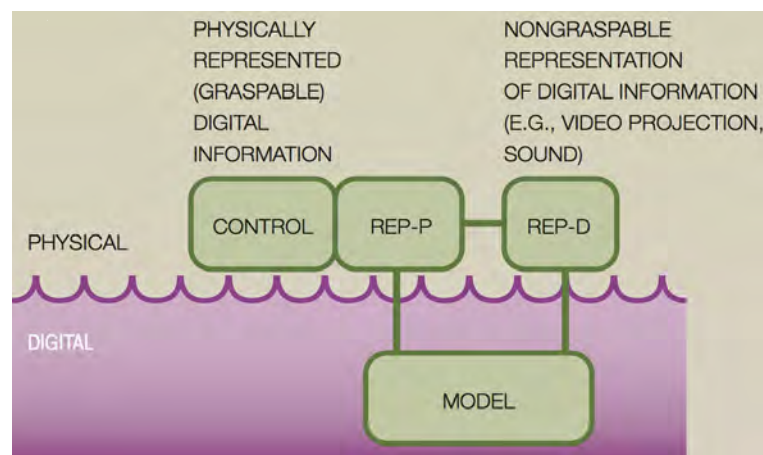


Figure 1.4: Interaction model of a TUI. Digital information is represented by a physical control. As regards IUIs, a dedicated device is represented by an everyday object.

The relationship between digital objects and their physical counterparts can be classified according to Koleva et al.'s coherence continuum.

Tangible User Interfaces can be classified regarding different criteria. One aspect taken into consideration is the degree of coherence, as proposed by Koleva et al. [2011]: "It is proposed that relationships between physical and digital objects can be rated along a coherence continuum, where the level of coherence represents the extent to which linked physical and digital objects might be perceived as being the same thing.". Figure 1.5 illustrates the coherence continuum for tangible interfaces and their physical objects. General purpose objects weakly resemble the virtual counterpart, whereas such physical objects that resemble the virtual counterpart strongest create the illusion of identity, i.e., the user perceives the physical object as representation of the virtual item. Instant User Interfaces do not limit the degree of coherence. Generally, any object replaces another (dedicated) object. The IUI concept leaves it to the user whether to appropriate an identical looking object as replacement or to choose a completely different one, as long as the affordances match. Like TUIs, IUIs also

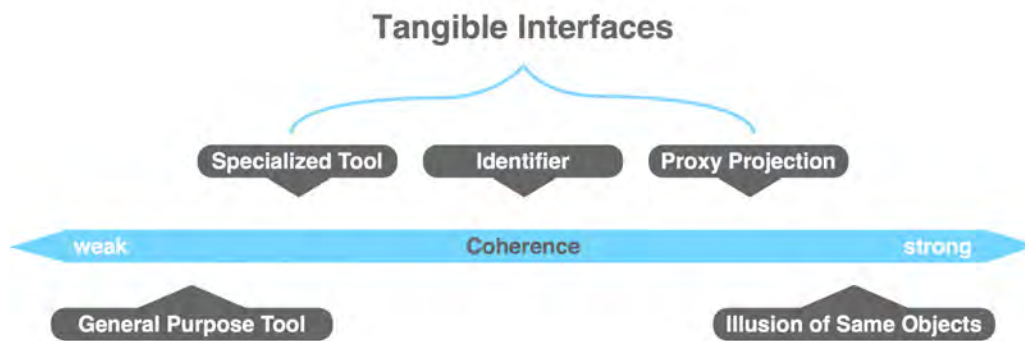


Figure 1.5: Kuleva et al.'s coherence continuum for physical objects and their virtual counterparts. Whereas TUIs provide various coherence between digital information and physical counterpart, UIs do so for dedicated objects and appropriated artifacts serving as their counterparts.

take on physical objects letting the user interact with the physical world in order to manipulate the virtual one. In this paradigm, any object serves as representation for a dedicated tool; ideally it mimics the original counterpart as to shape, weight and haptic properties.

However, TUIs principally focus on rigid objects as controllers. Objects with natural shapes, such as malleable surfaces, are more likely considered in *Organic User Interfaces* (OUIs).

1.3.2 Organic User Interfaces (OUIs)

Organic User Interfaces (OUIs) go one step further and add – apart from merely physical *input* as intended by TUIs [Holman and Vertegaal, 2008] – physical output to objects. The basic idea behind OUIs is that “an Organic User Interface is a computer interface that uses a *non-planar display* as a primary means of output as well as input. When flexible, OUIs have the ability to become the data on display through deformation, either via manipulation or actuation. Their fluid physics-based graphics are shaped through multi-touch and bi-manual gestures.” [Holman and Vertegaal, 2008] (emphasis added).

OUIs feature both input and output on non-planar object surfaces.

IUIs adopt interaction on non-planar surfaces from OUIs.

Instant User Interfaces do not explicitly focus on superimposing output *on* the object surface. However, what is adopted from the field of OUIs, is the physical richness of malleable artifacts in contrast with common interface design: “[the] [...] planar rigidity in interface design generally limits the usability of our computers in terms of their possible affordances.” [Holman and Vertegaal, 2008]. Like OUIs, IUIs want to fully exploit the physical and haptic properties of *any* object, thus also appropriate malleable objects of either plastic or elastic nature for providing the user a richer experience as deformable surfaces are perceptible by a user’s hand. Exploiting this property enables objects appropriated to the IUI metaphor to be used as haptic controllers allowing eyes-free interaction.

IUIs aim at hiding technology through interaction with everyday objects.

Bringing natural, organic shapes to interaction design, OUIs head for hiding technology. Vertegaal [2011] states: “The best thing that could happen to user interface design, then, is for computers to stop being technological devices and become just like real *everyday things* [...]” (emphasis added). This vision also attributes to IUIs and is originated from *Ubiquitous Computing* (UbiComp).

1.3.3 Ubiquitous Computing (UbiComp)

UbiComp targets at making computer technology disappear.

Introduced in the early 90s by Weiser [1991], Ubiquitous Computing targets at making computer technology disappear in such a fashion that it is not perceived by the user anymore. As the term “ubiquitous” suggests, UbiComp deals with everyday and omnipresent technology, being natural for the user such that she does not consider it a piece of special hardware or technology. In terms of IUIs, this metaphor targets at using *everyday*, hence ubiquitous artifacts people are surrounded by for appropriation, i.e., to turn these passive items into interactive controllers. This is the basis which all interface metaphors, i.e., TUIs, OUIs, and IUIs, have in common. What the concept of IUIs adds, however, is the idea of spontaneity, hence the provision of instant and ad hoc transformation of any object around the user to be exploited to her own needs. Ideally, this universal approach does not need an object to be specially prepared for interaction as this is the case for TUIs.

Having introduced the concept of IUIs alongside with its motivation and benefits, as well as its scientific background, we will now focus on the research questions towards IUIs that are investigated in this thesis.

1.4 Research Questions

In general, the work of this thesis investigates the appropriation of everyday objects by exploiting physical affordances of the artifacts for using them as haptic controllers and temporary substitutes to dedicated input devices for interactive systems. In more detail, the work deals with the following research questions:

- Which objects are within people's reach? In a diary study, we investigate people's work and leisure environments for understanding which objects come into question when considering appropriation of objects in the sense of IUIs.
- How do people actually appropriate these everyday objects (obtained from the diary study)? Based on a combination of Wizard of Oz (WOZ) and guessability study [Wobbrock et al., 2009], which presents users a certain outcome and asks them how they think it could be achieved, we present various scenarios in which a dedicated input device is necessary but not at hand. We target at identifying patterns regarding object types used and gesture types chosen by the testers to appropriate the missing controller. Which general findings can be extracted for the purpose of allowing spontaneous appropriation conform to the IUI metaphor?

This thesis investigates the appropriation of everyday objects as haptic controllers for digital interfaces.

We investigate everyday objects in a diary study.

We investigate appropriation patterns applied by users.

Regarding interaction with everyday objects, one challenge will be the distinction between voluntary system control with an object (thus, appropriation) and normal everyday object usage, thus non-voluntary system control. Therefore, another research question is:

We investigate the interplay of normal object usage and appropriation.

- Having a working system with real-time feedback available that features appropriation of everyday objects, how can voluntary system use be distinguished from non-voluntary system usage and how does this affect user experience?

We investigate spontaneity of access to IUIs.

- Furthermore, we want to find out whether using pre-defined gestures based on the patterns extracted from the WOZ study can be used to support spontaneity for IUIs for enabling instant appropriation. Is this a recommendable approach or does it influence the user experience for the worse?

This thesis focuses on user experience and not on technical feasibility of IUIs.

In this context we want to point out that we are interested in the user experience towards appropriation of IUIs and the resulting user experience. Although we present an interactive but limited prototype (cf. 5—“Testing Everyday Object Interaction with an Interactive Prototype”), we are not interested in the technical feasibility of IUIs, including object identification and tracking. We are certain that this technical challenge is a traceable problem, as recent advances in HCI and computer vision show, such as the KinectFusion system by Izadi et al. [2011] which supports robust and real-time 3D scene reconstruction on consumer hardware. In particular, our WOZ study assumes the existence of such technology.

1.5 Outline

The thesis is organized as follows:

- **Chapter 2.** We present a design space for everyday object interaction with digital interfaces. Various research projects listed in the design space related to everyday object interaction are discussed and contrasted with our work.
- **Chapter 3.** With respect to the first research question, this chapter presents a diary study conducted for finding out what kind of objects are within peo-

ple's reach. A quantitative analysis on the objects observed is followed-up.

- **Chapter 4.** With respect to the second research question, we present a WOZ study based on the objects identified in the diary study. This chapter describes the study and presents its results including a pattern analysis regarding which objects are manipulated in which way for a given set of scenarios.
- **Chapter 5.** This chapter describes the implementation of an interactive prototype allowing people to turn passive objects into active controllers. In a qualitative study, everyday object interaction based on the interactive prototype is tested with users. This also includes investigation regarding the discrimination between voluntary and non-voluntary system use.
- **Chapter 6.** We conclude the findings of this work and hint at future perspectives.
- **Chapter 7.** Having analyzed all research questions, this chapter provides guidelines for everyday object interaction, based on the findings from the previous user studies.

Chapter 2

Related Work

The idea of ad-hoc appropriation of passive ubiquitous things for breathing new life into them has drawn interest by researchers for the past five years. Early prototypes, such as presented by Carvey et al. [2006] and Mistry et al. [2008], started using everyday objects as tags for accessing digital content. In this chapter we describe current advancements in research that turn passive objects into interactive controllers or user interfaces.

Research interest in everyday object interaction draws more and more attention.

In order to summarize the various directions these research projects target at, we created a three-dimensional design space which examines the following aspects:

We created a three-dimensional design space for related work.

1. **Kind of Manipulation**, i.e., whether interaction with the object resembles *direct* or *indirect manipulation*.
2. **Types of Gesture**, i.e., which gestures the user can apply *on* an object (stationary) or *with* an object (non-stationary).
3. **Exploitation of Physical Affordances**, i.e., to what extent object-specific properties perceptible to the user are utilized.

Figure 2.1 shows the three-dimensional design space broken down into a printable model with overlapping axes. Each project is indicated by a set of connected squares. In

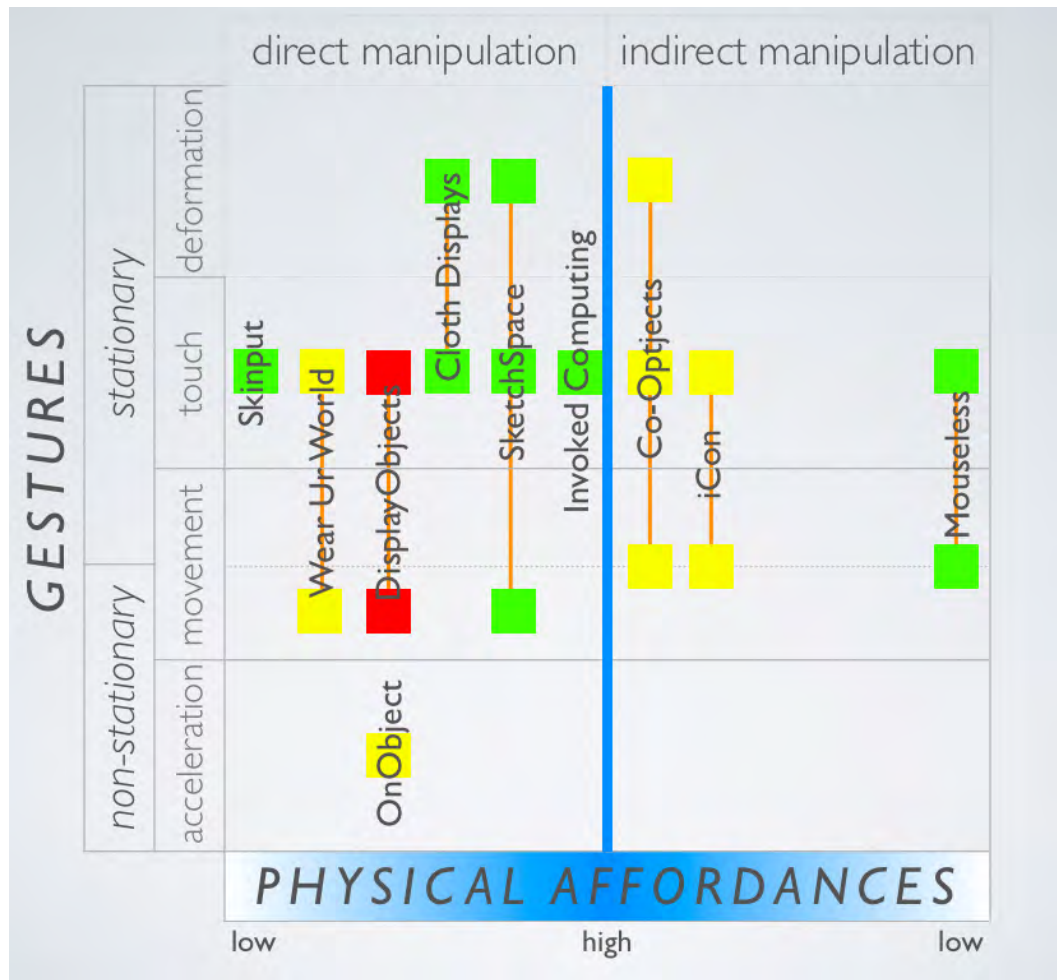


Figure 2.1: A design space for related work. The research projects are classified regarding kind of manipulation (x-axis, top), types of supported gesture (y-axis, left), exploitation of physical affordances (x-axis, bottom), and intensity of object preparation (green: low, yellow: medium, red: high).

Depending on the technology used, objects need to be prepared before they can be appropriated.

addition, the design space has regard to a fourth dimension, called “object preparation”. This is a rather technical attribute indicating the extent to which objects must be physically augmented before they can be appropriated. Green color symbolizes that no preparation is needed. Yellow indicates little to medium intensity, meaning that the objects have to be equipped with special markers. Finally, red means that additional effort is needed.

2.1 Design Space

Before describing each research project listed in the design space, we explain the interpretation of each of the design space criteria in detail.

2.1.1 Kind of Manipulation

In terms of augmentation of an (everyday) object, there is one basic question regarding which role the object plays. Two different perspectives come to mind here: *direct* vs. *indirect manipulation*.

Direct Manipulation

As regards direct manipulation, the object is to be considered a complete device substituting another. This means that both input and output are transferred to the everyday artifact. The object is the only thing that the user interacts with; there is no remote interface involved. In fact, input and output are superimposed on the object.

As regards direct manipulation, both input and output are superimposed onto the object surface.

DisplayObjects by Akaoka et al. [2010] is a good example based on direct manipulation. This system enables designers to decorate the surface of a physical prototype with visual styles and virtual UI widgets. The authors use two projectors for superimposing the output directly onto the object surface. Besides, simple interaction with the widgets is possible by multi-touch: Input is also superimposed onto the object surface. For DisplayObjects, both input and output share the same “location”, hence this example follows the design principle *input equals output*. The user does not need to change to focus between input and output; her locus of attention is fixed to the object surface. As the user can influence the output by directly manipulating the exact space where it is visualized¹, we deal with *direct manipulation*. Projects that make use of this

As a consequence, input equals output.

¹Basically, this is not constrained to the visual sense.

design principle are depicted on the left hand side of the design space (cf. Figure 2.1).

Indirect Manipulation

As regards indirect manipulation, input and output are distinct.

In contrast to direct manipulation, *indirect manipulation* is not compliant to “input equals output”. Effectively, input and output are perceptible at different “locations”. Take iCon by Cheng et al. [2010], for instance. The authors use everyday objects as controllers in order to manipulate a remote system, for example a desktop application. Putting up a cup and setting it down simulates a button click. However, this button is not available on the object, but geometrically distinct: The user manipulates the GUI visible on the computer screen by interacting with a cup – one device indirectly controls another one. As a consequence, the user must change the focus from input device to output device; the locus of attention changes.

In order to be able to control existing systems, our work focuses on indirect manipulation.

The work presented in this thesis follows this principle. Co-Objects are remote controllers made of everyday objects with which the user interacts with her hand in order to manipulate an existing remote target system. Related work that conforms to the principle of indirect manipulation is visualized on the right hand side of the design space (cf. Figure 2.1).

2.1.2 Types of Gesture

The design space regards four different types of gesture.

One important aspect of consideration is how the user can interact with an everyday artifact, i.e., which types of gesture are supported for manipulation. With respect to related work, we classified the supported gestures into four major categories: touch, deformation, movement, and acceleration. More roughly, touch and deformation are grouped into so-called stationary (upper half of the design space, cf. Figure 2.1) and non-stationary (lower half of the design space, cf. Figure 2.1) gestures. Stationary gestures are applied locally *on* the object. Non-stationary gestures basically involve the manipulation of the object

as a whole. Touch and deformation are selective, thus stationary. Acceleration is applied on the entire object body, hence non-stationary.

Movement is basically non-stationary. However, what about turning a bottle cap, for instance? Certainly, the cap is (re-)moved, thus motion is involved. Nevertheless, the main part of the object body rests where it is. Consequently, we attribute movement to be both stationary and non-stationary. Furthermore, deformation can be considered a special variety of (pressure-sensitive) touch. Similarly, acceleration is a two- or three-dimensional motion gesture adding information about velocity for each point in time.

Motion can also be applied *on* the object surface.

Stationary Gestures

Stationary gestures are gestures the user applies locally *on* a specific part of the object.

- **Touch.** By touching different regions on the artifact surface using one (single-touch) or multiple (multi-touch) fingers at the same time, the user can cause events. A virtual button, for instance, is a typical example for an interface dependent on touch. A physical button, on the contrary, requires the user to put some pressure onto the control. Such gestures belong to the category of deformation.
- **Deformation.** Deformation can be considered an extreme form of touch. By putting pressure onto a particular region on the object surface (e.g., with the finger), a user can change the shape of an artifact. Depending on the object material, the artifact can return to its initial shape after pressure is released (elastic), or it can keep the current form even when the user removes pressure (malleable). A physical button, for instance, is considered elastic, as pressure temporarily changes the overall object shape.

Objects can react to touch events.

Deformation combines both touch and pressure.

Non-stationary Gestures

Gesture types belonging to this category imply planar or spatial movement of the interaction artifact. Basically, interaction *with* the entire object is of interest rather than considering special areas of its surface.

Objects can be moved.

- **Movement.** Gestures caused by planar or spatial movement applied to the interaction artifact belong to this category. Actuation, i.e., movement initiated by the object itself, is not in our focus, as everyday objects should be altered as little as possible (and in general, everyday objects do not move). As mentioned before, turning, for instance, can be both non-stationary and stationary. The latter applies when only a special part of the object is concerned, e.g., when turning a bottle cap.

Objects can be accelerated.

- **Acceleration.** Acceleration gestures also result in movement of the interaction artifact by also taking temporary speed an object is moved at into consideration. Since such gestures can be recorded and recognized by an accelerometer, we classify them in the category of acceleration.

2.1.3 Exploitation of Physical Affordances

Objects are distinguished by physical properties.

A physical object has different physical properties, making the object distinguishable from another one when the properties are different. Such properties are of either physical or chemical origin, e.g., due to chemical composition of the materials the object is made of. Shape, weight, elasticity, or surface texture are examples for such specific properties.

Physical affordances suggest how an object can be manipulated.

Alongside with these physical attributes, a user often anticipates a certain idea of manipulation, i.e., by looking at the object, by feeling it out etc., the user anticipates how the object could be used and which gestures she could apply to the object. A ball, for instance, suggests to be thrown away and to be caught. This principle of anticipated manipulation is denoted by *Physical Affordances* [Norman, 2002].

Since everyday objects come in various shapes, weights, colors etc., they also embody various physical affordances. This richness of affordances is an ideal component to be exploited for everyday object interaction, at least when using these quotidian artifacts as temporary controllers. As regards Co-Optjects, for instance, physical controls-like add-ons that decorate an object (e.g., a bottle cap) can be used for appropriation. Consequently, the interface can be controlled purely manually and without having eye contact as the user can feel the cap itself and therefore *where* it is located on the object.

Hence, exploiting physical affordances brings more comfort to interaction. A physical button can be perceived using either the haptic or visual input channel. A virtual button on a flat surface, to the contrary, can only be perceived by vision. The more a research project listed in the design space exploits physical object affordances for interaction, the closer it is oriented towards the blue center line. Co-Optjects, for example, strongly make use of this concept, which allows everyday objects to be used as haptic controllers, i.e., without having to look at them permanently while interacting with them.

Co-Optjects exploit haptic object properties.

2.1.4 Co-Optjects in the Design Space

Co-Optjects target at the appropriation of everyday objects as controllers by exploiting physical affordances. Therefore, it is listed closely to the right of the center line. Moreover, Co-Optjects include both stationary (touch and deformation) and non-stationary (planar movement and turning) gestures. As objects are partially dependent on paper fiducials, object preparation is denoted by little/medium.

Co-Optjects exploit physical affordances, feature indirect manipulation, support various types of gesture, and need medium object preparation.

2.2 Related Research Projects

In this section we describe all research projects listed in the design space and contrast them with our work. We give a brief description, name the benefits and contrast

Each research project listed in the design space will be described in detail.

each project with our work. We start with Mouseless and Skinput as these somehow define the border of our design space: They eliminate the need for dedicated input devices. The Remaining projects are presented subsequently. The closer a work is related to ours, the later it is listed.

2.2.1 Mouseless: Mistry and Maes, 2011

Mouseless features computer mouse-like interaction by synchronizing a user's hand movement with a mouse cursor.

Motivated by reducing the form factor of dedicated input devices – such as the computer mouse – in order to enhance mobility, Mistry and Maes [2011] presented a ubiquitous system that maps a user's hand movement on a flat surface to a mouse cursor on a screen. The user pretends to be holding a physical mouse underneath the palm, which the system interprets as real mouse movement. Unlike previous attempts in research and industry trying to minimize form factors of pointing devices, Mouseless contributes to completely eliminating the controller physics, whilst maintaining a feeling close to reality. Besides common mouse gestures, Mistry and Maes add recognition of multi-touch gestures, such as pinch-to-zoom.

When the user moves her palm, emitted IR light is reflected and captured by a camera.

Technically speaking, Mouseless requires an IR laser and a camera with an IR pass filter which can be embedded in the side of a laptop bottom case or a keyboard, as depicted in Figure 2.2. The laser creates an invisible light plane on the surface next to the computer. By putting the hand in a curved position on the surface, the IR light is reflected by the flesh and captured by the camera. Using common thresholding techniques from computer vision research, these blobs representing the user's hand are extracted and the mouse cursor is synchronized with the movement of the hand (i.e., the detected blobs). Similarly, vertical finger movement is interpreted by the system as a mouse click. Using this technique, Mouseless detects left and right click, scrolling by palm curling, and zooming via pinching. Adding a second laser on top of the first one creates a second plane for detecting two-level gestures such as drag and drop by pinching thumb and forefinger, lifting the hand up, and putting it down again elsewhere. In order to support natural feel of a physical mouse, clicks are augmented through audio effects.

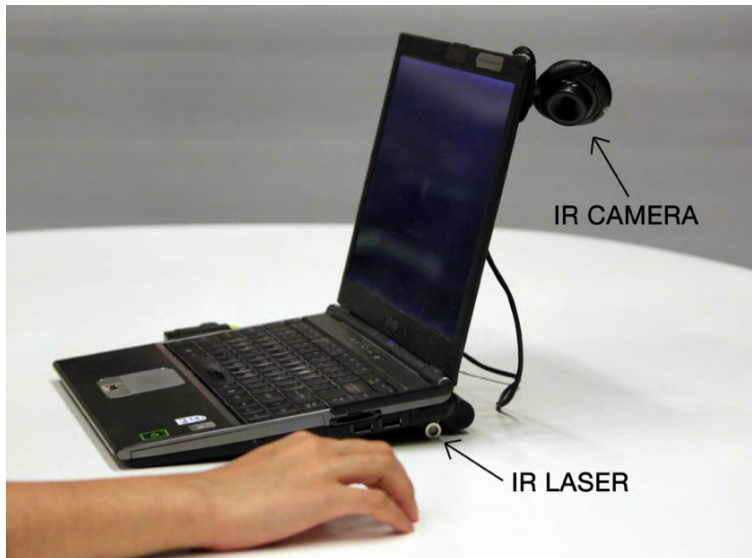


Figure 2.2: Physical setup for Mouseless. An IR laser and a camera are integrated to the side of a laptop. When the user moves her palm, IR light is reflected and captured by the camera. Computer vision algorithms help to extract spots and synchronize their movement with the mouse cursor.

The benefit of Mouseless is an almost natural feel of mouse control without having to carry along a physical device. Like Co-Optjects, Mouseless is independent of a dedicated piece of hardware by exploiting ubiquity of everyday “objects” which is in this case a human hand. Yet, there are differences between both projects:

Mouseless is independent of a dedicated piece of hardware.

- Mouseless targets at eliminating physics and therefore also physical affordances which in turn are exploited by Co-Optjects. Mouseless uses *any* flat surface in the environment, whereas Co-Optjects focus on *particular* objects appropriate to the task to be executed.
- The intentions of both projects are different. While Mouseless intends to eliminate *permanently* the physical mouse, Co-Optjects claim only *temporary* use of non-dedicated input devices.

- Finally, Mouseless features a lot of *multi-touch* gestures, whereas Co-Optjects – in terms of touch detection – consider *pressure-sensitive touch*.

2.2.2 Skinput: *Harrison et al., 2011*

As mobile devices, such as smartphones, merely provide a small input and output space to the disfavor of a user's interaction comfort, Harrison et al. [2010] investigated the human's skin – with its two square meters of external surface area – as large and omnipresent interaction surface.

Skinput is capable of detecting and locating touch on the human's skin using vibration sensors.

Responsiveness to the human's skin was achieved by exploiting the fact that touch causes flesh and bones underneath the skin to vibrate. Harrison et al. constructed an armband with two arrays of five vibration sensors each made of piezo films, as depicted in Figure 2.3. The highly sensitive instruments capture acoustic waves which vary in their frequency depending on *where* the touch triggered the vibration. This makes multiple touches distinguishable from each other by their geometric origin on the skin. The captured audio signal is processed by a client which is finally connected to a visualization application over a socket connection. As the wave form varies depending on the user (variety in the proportion of fatty tissues) and the position of the armband, Skinput is dynamically calibrated by a machine learning algorithm. The user trains Skinput by tapping multiple times at the skin regions of interest.

Skinput allows the user's arm to be exploited as mobile interaction space.

User studies have proven Skinput to detect true positives by 88,8% for six distinguishable touch points on the forearm, wearing the sensor above the elbow (cf. Figure 2.4). Even for body movement, e.g., jogging, the accuracy does not significantly decrease. Furthermore, touch recognition has been tested with small interfaces projected on the user's skin by adding a pico projector to the armband. Skinput is further able to be trained to detect surfaces and materials, such as a table, for instance. Touching the artifact causes the wave signal to change specifically, for Skinput is able to tell which object has been tapped.



Figure 2.3: The Skinput prototype armband. Two arrays consisting of five vibration sensors each capture vibration of flesh and bones caused by touch.

The overall benefit is its ubiquitous approach of appropriating the human's skin for a touch-sensitive interaction area. Like Co-Optjects, Skinput exploits ubiquity (in this context: of the skin) to become less dependent on dedicated input devices drastically limited in size, such as multi-touch screens used in smartphones. Significant differences between both projects are:

Skinput exploits the ubiquity of the human's skin.

- Skinput focuses on general augmentation of any skin surface, which is less specific compared to Co-Optjects being targeted at diverging objects.
- Unlike many physical objects, skin is a relatively flat surface, thus Skinput does not exploit variety in haptics for interaction. Co-Optjects, however, do.
- Finally, both projects take eyes-free interaction into consideration. Co-Optjects do so by guiding the user through feeling out the specific object surface, whereas for Skinput the user remembers the touch location on her own skin.



Figure 2.4: Input accuracy achieved by Skinput. Worn above the elbow, the sensor can distinguish six touch zones on the user's forearm featuring a touch detection of true positives by 88,8%.

2.2.3 WUW – Wear Ur World: Mistry *et al.*, 2009

Unlike WUW, information displays are often locally separated from the item information is requested for.

Mistry *et al.* [2009] investigated gaining instant access to information, for example about (everyday) objects and products, regardless of location and time. Heretofore existing systems – the authors argue – are neither instant nor is the information directly related to the product with the locus of attention concentrated on external devices/displays, such as a smartphone, for obtaining information. A sample scenario could be a client in an electronics shop looking for user ratings for a particular vacuum cleaner he is interested in buying. Therefore, he pulls out his iPhone and searches for the ratings on the internet.

WUW projects product information onto the object surface the user can interact with by applying freehand gestures.

WUW (Wear Ur World) bridges the local and logical gap between object and information display using a camera and projector prototype attached to a cap the user is supposed to wear on his head, as depicted in Figure 2.5. By visually tracking fiducials that have to be put on the object of interest, the system can both detect and track the object the user is looking at while the projector displays the product information directly onto the artifact. Using finger markers, the user can also interact with the projected interface through multi-touch gestures, such as zooming and panning. Besides, WUW features freehand gestures, such as forming a frame with forefinger and thumb (cf. Figure 2.5) for telling the web camera to take pictures. In addition, iconic gestures such as drawing the “@”-sign for launching an email application, are understood.

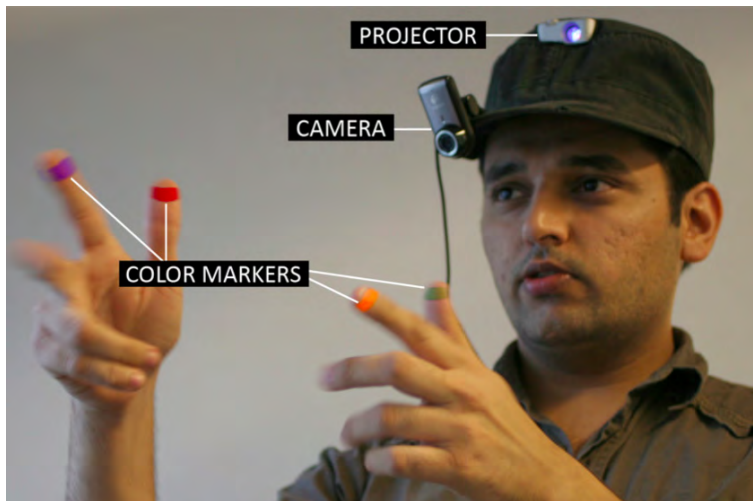


Figure 2.5: A user wearing the WUW prototype on his head. A projector projects a virtual user interface onto any surface. The camera tracks the movement of color markers for interpreting freehand gestures.

Recalling the example about the client, imagine he were using WUW, he would simply look at the vacuum cleaner and the projector would immediately display related information *on* this product, such as user ratings collected from the internet. The man could simply touch on the virtual “Show Ratings”-button which would present a list with the corresponding results. Hence, the product and its information are *directly* linked: The product itself becomes the interactive information display.

Thus, WUW enables passive objects to become interactive ones. The benefit of WUW is its way of linking visual information instantly to the object concerned through on-surface projection and interaction. Both Co-Optjects and WUW have in common that they turn passive (everyday) objects into interactive gadgets for the user. Yet, there are significant differences between both research projects which are as follows:

- WUW is rather an *information display* for objects, regardless of surface shape or texture. Co-Optjects,

As an example, WUW can be used to obtain user ratings for a product.

WUW turns passive objects into interactive interfaces.

however, use objects as *controllers*. There is no information to be displayed on a Co-Optject.

- Co-Optjects strongly focus on physical affordances of an object, being a tactile controller for *indirect manipulation* targeted at enabling eyes-free interaction. WUW does not exploit haptics of the object at all but makes use of *direct manipulation* and freehand gestures. As information is visualized, eyes-free interaction is not regarded.

2.2.4 Cloth Displays: *Lepinski and Vertegaal, 2011*

Cloth Displays is a flexible piece of cloth that turns any object into an interactive surface.

Unlike traditional user interfaces which are dedicatedly designed for rigid surfaces, such as liquid crystal displays, Cloth Displays developed by Lepinski and Vertegaal [2011] is a flexible and lightweight approach bringing user interfaces to any object surface. By draping an object with this interactive piece of cloth, a tight coupling between object and interface is established, as the interface follows the organic object shape, as demonstrated in Figure 2.6. With this metaphor, “affordances provided by the object’s shape become part of the available interaction vocabulary” [Lepinski and Vertegaal, 2011]. Cloth is a material well known by humans as is how it can be manipulated: It can be folded, stretched, and it adopts dynamically the shape underneath it.

Technically, Cloth Displays requires a fiducial tracking system.

For making cloth interactive, a tracking system similar to the [Vicon Motion Capture System](http://www.vicon.com)² is used to track the fiducials fixed to the cloth and a user’s fingers. Using a ceiling-mounted projector, the cloth also serves as projection surface for visual interfaces (cf. Figure 2.6).

A physics engine is used to closely approximate the physical behavior of cloth.

On the software side, the researchers implemented a physics engine which calculates a surface model that closely approximates the physical behavior of the cloth to be consistent with the real world counterpart: “this model deforms naturally, can fold over itself, and can be stretched, returning to its original shape once released.”.

²www.vicon.com

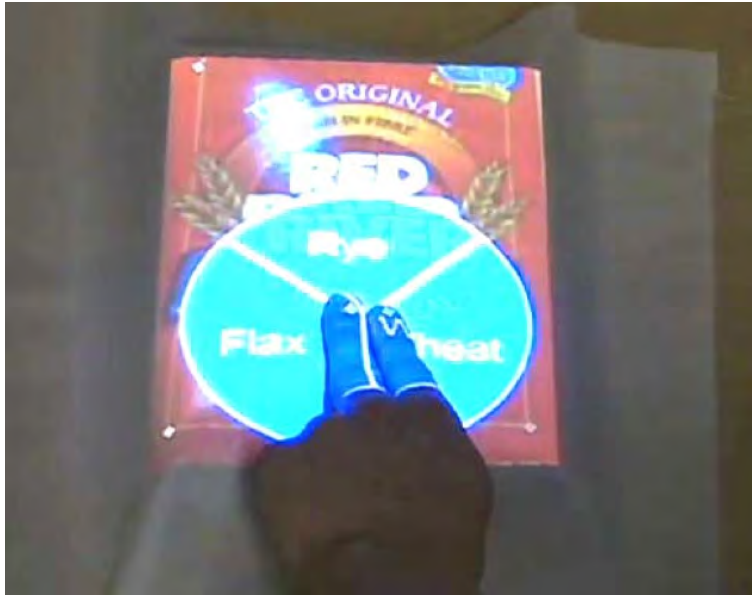


Figure 2.6: The Cloth Displays prototype. As soon as the piece of cloth drapes an object, it adopts the artifact shape and provides a touch-sensitive GUI.

Basically, Cloth Displays follows three major design principles:

1. **Metaphorical Compliance** with cloth through the physics engine
2. **Input Equals Output** by making the projected interfaces on the cloth touch-sensitive
3. **Form Follows Function** as the cloth adopts to the underlying object shape

In accordance with these principles, the user can drape objects, stretch, squeeze, pinch and peel, touch, and hover above the augmented piece of cloth.

The benefit of Cloth Displays is to instantly turn any surface into an interactive display adopting the surface shape. Besides, Cloth Displays is a stand-alone controller for remote interfaces via typical cloth manipulation, such

As regards manipulation, Cloth Displays features stretching and folding.

as stretching and folding. Reflecting the goals of the work presented in this thesis and Cloth Displays, both exploit (everyday) object shapes for interaction. The differences are identified as follows:

- The design principle *input equals output* manifests – similar to WUW – direct manipulation, using the object as touch-sensitive interface. Co-Optjects, on the contrary, are indirect haptic controllers, keeping the locus of attention at the device to be controlled.
- Finally, Co-Optjects target more intensively at haptic properties compared to Cloth Displays, as draped cloth eliminates surface-specific tactile properties.

2.2.5 OnObject: Chung et al., 2010

OnObject lets users link acceleration-based gestures to physical objects.

With OnObject, Chung et al. [2010] have taken a different approach regarding the appropriation of everyday objects, since they combine acceleration-based gestures with audible output linked to objects. Visual output, such as on-surface projection, is not addressed in this work. The major goal of OnObject is “to provide a way for novice end users to rapidly transform physical interfaces to gesture interfaces to their liking in the very context of use, without a lengthy development process.” [Chung et al., 2010].

OnObject consists of an RFID reader, an accelerometer, a microphone, and a speaker.

Technically, OnObject consists of a portable RFID (Radio Frequency Identification) reader with an integrated microphone and speaker, as shown in Figure 2.7. Besides, the gadget can both capture and recognize gestures based on acceleration, such as shaking, swinging thrusting, tilting, and applying circular motion. For in-situ programming, the user simply holds the reader in her hand and grabs an object. Its RFID tag is immediately read. Then the reader captures the gesture the user applies on the object. Pushing the button on the gadget records the users voice which is then linked to the tag and the gesture applied before. Whenever the object is grabbed and the same gesture is executed again, the voice is played back. Hence, OnObject gives non-computer experts access to end

user programming on any physical object for making it sensitive to gestures of movement.



Figure 2.7: End user programming with OnObject. The user puts an RFID tag to the object (“P”) to be programmed and executes an acceleration-based gesture while holding the reader. It reads the tag, captures the gesture, and records the user’s voice.

One field of application for OnObject is e-learning. Users could program words on objects for learning a foreign language. By shaking the object, the RFID reader would tell how the word representing that object is to be pronounced correctly.

OnObject is used for e-learning.

The clear benefit of OnObject is thus the instant and easy linking of gestures and sound to objects. What both research projects, i.e., Co-Optjects and OnObject, have in common is the fact, that gestures can be applied on objects. Ultimately, both projects are very different:

OnObject features end user programming without requiring computer expertise.

- OnObject focuses on end-user programming. Co-Optject does not as it uses predefined gestures, as experienced in a WOZ study.
- Furthermore, Co-Optjects strongly concentrates on haptic properties of everyday objects for turning them into haptic interactive controllers. OnObject does not exploit the haptics of an object, as it recognizes acceleration-based gestures only.

- Principally, Co-Optjects are controllers that can be linked to any application. Objects programmed with OnObject are only linked to sound output so far, thus resembling an audible information display.

2.2.6 DisplayObjects: Akaoka et al., 2010

DisplayObjects
reduces the process
of time-consuming
physical prototyping.

In the field of prototyping, quickly changing the visual appearance of a mockup as well as adding interactive functionality is challenging, as it is often “difficult to smoothly integrate hardware controls like buttons or dials into a product’s display at this [i.e., an early] stage.” [Akaoka et al., 2010]. In order to reduce the time-consuming process of physical prototyping, Akaoka et al. developed DisplayObjects which lets the designer augment passive mockups or objects with on-surface projected visual styles and user interface widgets.

The user can drag
virtual UI widgets to
the physical
prototype by touch.

Similarly to a painter, the prototype designer holds a physical palette with projected UI widgets that can be added to the mockup by simultaneously touching the virtual widget and the physical prototype. This end user programming for non-experts paired with the creation of seamless interactive surfaces for *existing* prototypes generates the basic design requirements for DisplayObjects.

DisplayObjects
requires the user’s
fingers and the
prototype to be
equipped with
spherical markers.

In terms of system implementation, the research team used a Vicon tracking system: The prototype, the palette, and the user’s fingers must be fitted with spherical markers. Projectors are used as virtual display for the objects and the palette. Both must be spray-painted with reflective paint; otherwise, the projection would not be clearly visible. The DisplayObject framework matches the marker data received by the Vicon with a stored 3D model of the object. This way, the projection of the UI is synchronized with its target even when the object is moved spatially or rotated. The hardware setup is sketched in Figure 2.8.

Figure 2.9 demonstrates the ad-hoc programming process of a virtual keyboard on a naked styrofoam block turning it into a mobile phone prototype. The projector projects

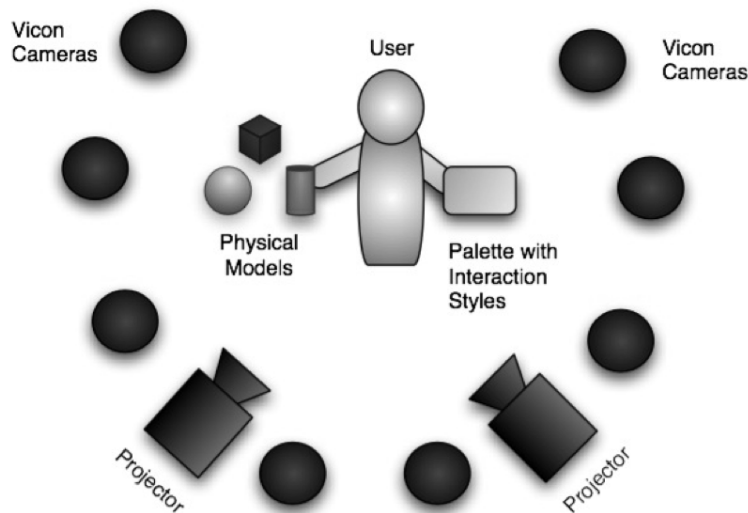


Figure 2.8: Hardware setup for DisplayObjects. The room needs to be equipped with multiple projectors and tracking cameras in order to track physical models and the interaction palette used as projection space for virtual UI widgets.

various UI widgets onto the paper tablet, which the user can tap with her finger by simultaneously tapping with the other hand on the target surface. This causes the UI element to be projected on the object. Finer adjustment is achieved through multi-touch gestures, such as pinch-to-zoom for resizing the virtual keyboard. Interactive scripts that react to single touch events on the mockup can be added but require programming expertise and a computer. So, basically, the prior intention behind DisplayObjects is to simplify and accelerate the process of visual redesign of the object surface of a mockup. In terms of appropriation of everyday objects, DisplayObjects was tested with students who created an interactive calculator interface projected on the backside of a credit card for transferring payments, as shown in Figure 2.10.

Adding visual style to passive objects requires no programming expertise.

The benefit of DisplayObjects is its easy-to-use graphical end user programming interface that allows non-computer experts to decorate mockups or everyday objects with a virtual user interface within minutes (physical object preparation with markers and spray paint being ignored). In comparison to Co-Optjects, both systems target at ex-

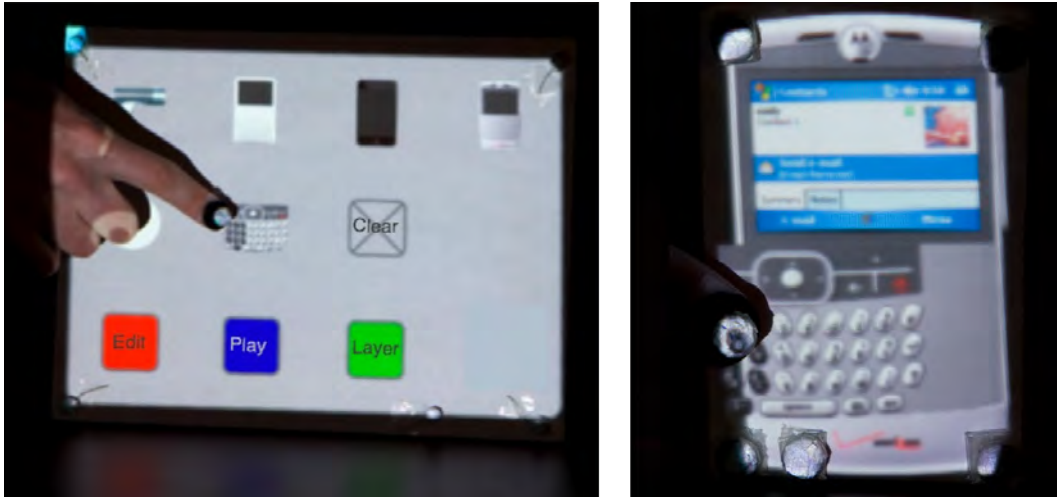


Figure 2.9: Palette with interaction styles (left) and physical mockup with a virtual interface (right). The user touches an UI widget on the palette and drags it to the naked prototype for adding the widget to it.

exploiting (everyday) objects for adding interactive behavior. The differences are as follows:

- DisplayObjects is targeted at prototyping and the augmentation of the look of device prototypes. Interaction is a minor field of interest and only works for single touch gestures. Co-Optjects are not concerned with visual object style and target at interaction through planar rotation, movement and pressure-sensitive touch on objects, making them instant controllers.
- Co-Optjects preserve and exploit physical affordances of an object. DisplayObjects alters the affordances with spherical markers and spray paint, which could change the tactile surface texture.
- Furthermore, DisplayObjects projects the UI directly onto the object, unlike Co-Optjects which make use of indirect manipulation.
- In addition, DisplayObjects addresses easy-to-use end user programming which is not used in the Co-Optjects prototype since pre-defined gestures are being tested.



Figure 2.10: An augmented credit card (left: front side, right: back side). Display-Objects has been used to add a graphical user interface (back side) for being able to transfer payments directly.

- Finally, Co-Optjects support the idea of spontaneous and natural interaction with everyday objects to larger extent than DisplayObjects does. The latter requires time-consuming setup and calibration of a tracking system, as well as object preparation such as spray painting (which is appropriate in the field of prototyping). Yet, interaction is unnatural as it requires the user to wear a trackable ring, unlike Co-Optjects.

2.2.7 SketchSpace: Holman and Benko, 2011

Similar to Akaoka et al. [2010], Holman and Benko [2011] argue that – in the field of prototyping – adding interactive behavior by embedding physical sensors in a mockup is a complex and time-consuming process. Testing for the best suitable sensor requires it to be physically embedded in the mockup without altering the shape. Often, ideas are not even tested as designers fear putting too much effort into a physical prototype which may be rejected later anyway³.

Designers consider physical prototyping a time-consuming process.

³e.g., due to interaction flaws detected in a user test

SketchSpace features rapid prototyping by replacing physical sensors with virtual ones.

Using SketchSpace, the process of sketching which is distinguished from prototyping by its intentional roughness [Buxton, 2007], can be dramatically simplified and accelerated. Instead of embedding physical sensors, a designer can rapidly equip a prototype or object with virtual sensors that imitate the physical equivalents, allowing the mockup to detect touch, grasp, spatial position, rotation, proximity of objects to each other, and deformation. This simplifies the testing of a plentitude of interaction variants without changing the mockup physics.

SketchSpace uses a combination of projector and depth camera for projecting information onto physical objects and for recognizing gestures.

Technically, SketchSpace is based on LightSpace developed by Wilson and Benko [2010]. A projector and a depth camera are mounted above the designer's workspace, as visualized in Figure 2.11. The depth camera is responsible for object tracking and for virtualizing the mentioned sensors, whereas the projector adds visual feedback to the passive object. With the aid of a software panel, the user can map the virtual sensors which are responsible for communicating physical state changes of the artifact to the system. For example, the designer can equip her artifact with a virtual button by determining its size and position on the object. This button can then be linked to any mouse or keyboard event for communicating with desktop applications. Whenever the virtual button is touched, the linked event is fired. Multi-touch gestures are not supported as these would interfere with grasp detection. Similar to the button, any virtual sensor can be added to the object of choice and linked with any event. As an example, using SketchSpace, a designer can test the interaction of bending the corner of a sheet of paper for navigation to the previously visited web browser page, as investigated in Holman et al. [2005].

SketchSpace enables designers to explore new input modalities for existing input devices.

The benefit of SketchSpace is its great flexibility in rapidly adding various virtual sensors to any object or material in order to test new input modalities for existing passive prototypes or devices, such as a tilt-enabled mouse (cf. Figure 2.12). Both SketchSpace and Co-Optjects have in common that objects can be turned into interactive controllers. The differences between both research projects are the following:

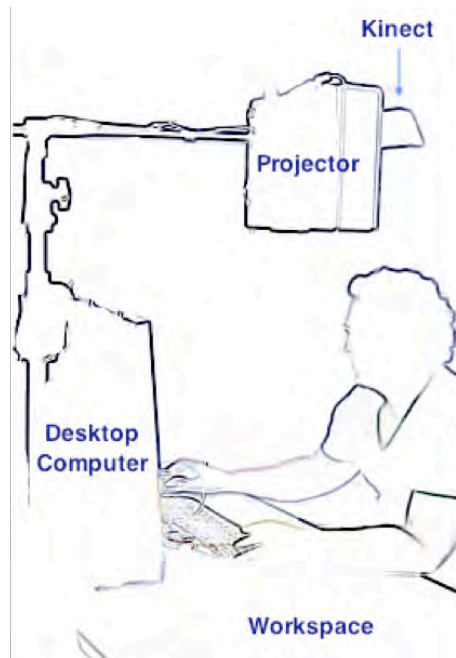


Figure 2.11: The SketchSpace hardware setup. A projector and a depth camera (Kinect) are mounted above the user's workspace. The projector projects visual information onto artifacts, whereas the depth camera interprets object manipulation.

- Intentionally, SketchSpace targets at sketching and rapid prototyping for new or existing input devices. Using virtual sensors disregards physical affordances, since implementing real sensors may alter weight, size and surface structure. Co-Optjects accept the physical design as given and exploit the affordances
- Furthermore, Co-Optjects are not solely targeted at desktop applications. Everyday tasks and interfaces, apart from the working environment, such as regulating ceiling light, are also regarded, unlike SketchSpace.
- From a technical point of view, SketchSpace offers a richer gesture repertoire compared to Co-Optjects. Yet, SketchSpace is limited in (a) the number of ob-

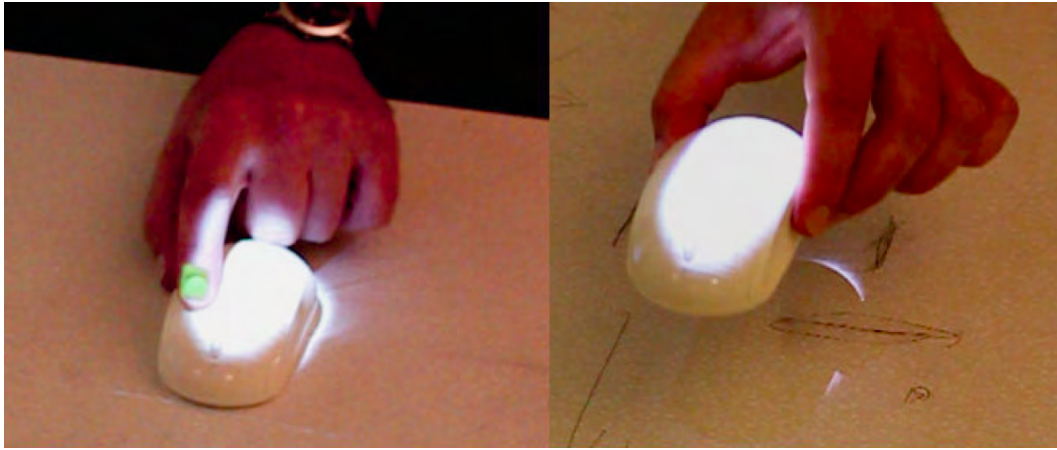


Figure 2.12: A tilt-enabled computer mouse. Using SketchSpace, a virtual tilt sensor has been added to an ordinary computer mouse, such that the mouse can be tilted in order to fire an event.

jects that can be simultaneously used⁴ and (b) touch- and pressure detection. SketchSpace only supports general object deformation but does not recognize deformation at distinct regions on the object surface. The Co-Objects prototype is capable of sensing different push states through pressure intensity, for example for a stapler that has two different push levels (cf. 5—“Testing Everyday Object Interaction with an Interactive Prototype”).

2.2.8 iCon: Cheng *et al.*, 2010

People arrange everyday objects spatially on their desk when arriving at work.

Setting their sights at enhancing desktop work productivity, Cheng *et al.* [2010] investigated using everyday objects people have at their office as instant controllers for quickly controlling background tasks. Having learned that people use their spatial memory for arranging similar objects at hand’s reach on their desk when arriving at the office, the authors concluded that these objects are suitable to be used as auxiliary tabletop input devices in work environments.

⁴SketchSpace allows to manipulate two objects at the same time at maximum.

Inspired by these results, Cheng et al. implemented iCon, a system that allows ad-hoc transformation of personal everyday objects into controllers for desktop applications. Two prototypes were developed which use standard web cameras for tracking paper fiducials put on the everyday object. With the aid of the reacTIVision framework developed by Kaltenbrunner [2009], iCon can recognize binary gestures such as patting an object or lifting it up and down and consecutive input such as planar rotation and movement applied on an everyday object. Regarding the first prototype, the camera is positioned above the user's desk, having a bird's eye view on the objects put in the workspace in order to detect tapping on objects with a closed top surface. For detecting the lift-up-and-down gesture applied to an artifact with an open top surface, such as a cup, the second prototype uses a camera mounted beneath the desk. However, for this to function, the table surface has to be transparent, similar to DSI- (Diffused Surface Illumination) [Akechi et al., 2011] or FTIR- (Frustrated Total Internal Reflection) [Han, 2005] based multi-touch tables just without IR-LEDs, as the fiducials are placed underneath the object.

iCon uses paper-based fiducials put on the objects for gesture detection.

In-situ linking of objects with gestures and executed actions is done via the iCon user interface (cf. Figure 2.13). The user puts a fiducial sticker on an object and puts it in the camera periphery. It appears immediately in the iCon programming interface where the user can link it to predefined actions of different applications via drag and drop.

Passive everyday objects are linked to desktop application commands via drag and drop.

In order to distinguish between normal object usage and intended appropriation as desktop controller, iCon uses a time-to-live mechanism. A click, for instance, is mimicked by lifting an object up and down again. If the user puts the object down within a short period of time, the click is fired. In other cases, e.g., whilst drinking which means, that the user holds the cup for a longer timer, the event is cancelled.

iCon distinguishes appropriation from normal object usage based on a time-to-live mechanism.

The enhancement of work productivity using iCon as controllers has been confirmed by a quantitative study. On average, testers significantly eliminated 6.67% of interruption time while working on a document, as they transferred the control of typical tasks, such as altering the

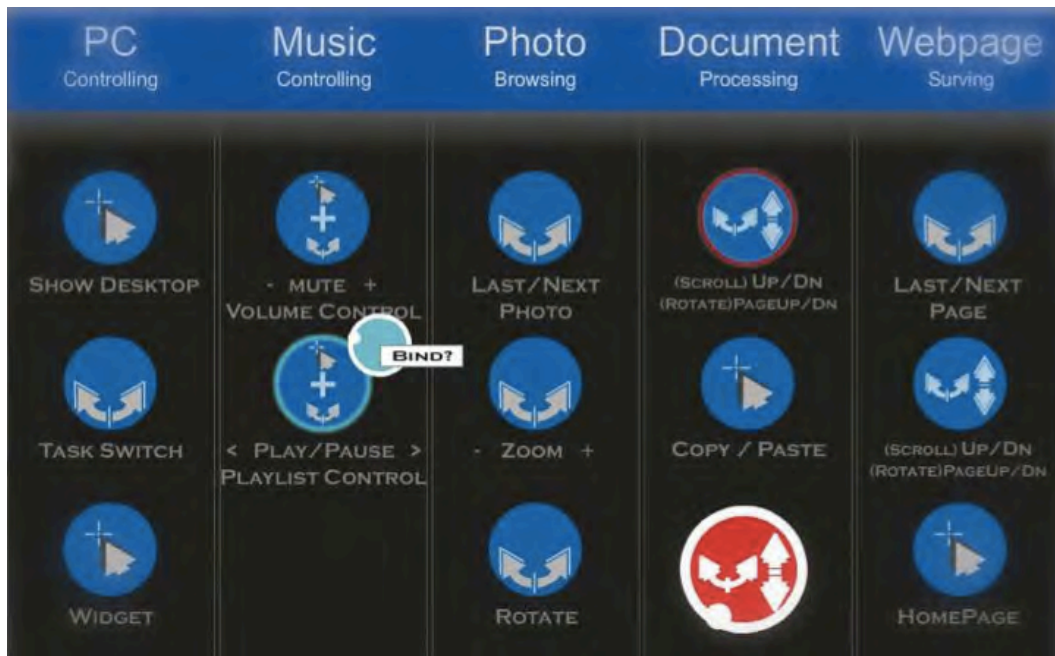


Figure 2.13: The iCon end user programming interface. As soon as an iCon is detected in the workspace, its virtual counterpart appears on the screen (red icon). Using a computer mouse, the user can bind the physical widget to an application command via drag and drop.

iCon significantly enhances work productivity.

music volume, to everyday objects, such that they did not need to toggle focus between word processing application and music player application. In a subsequent qualitative study testers stated, that everyday objects are useful controllers for low precise input with low engagement and medium-high frequency of usage, partially without having to look at the object while appropriating it.

The main benefit of iCon is, that it allows users to program everyday objects to their needs which saves them focus switching time between different main and minor desktop applications in use. Both Co-Optjects and iCon target at using everyday objects as controllers for indirect manipulation of user interfaces. Yet, there are differences:

- Co-Optjects address spontaneity in a stronger sense by making use of predefined application-object mappings. In contrast, iCon features a higher flexibility as the user can freely bind objects to applications. Yet,

this system is dependent on a dedicated transparent table.

- iCon does not regard deformable object properties and therefore loses important haptic features of everyday objects. A click imitated by a lift-up-and-down gesture which was rarely observed in our WOZ study (cf. 4—“Wizard of Oz Study: Interaction Patterns for Instant User Interfaces”) is less natural than a push gesture. The latter is ideally supported by Co-Optjects, since deformable object properties are exploited.
- Unlike iCon, Co-Optjects also investigate the usage of everyday objects as controllers in leisure scenarios.

2.3 A Futuristic Perspective on Everyday Object Interaction

The following work also attributes to the appropriation of everyday objects by turning them into interactive devices. Although it is still a work in progress, it is not of minor interest.

2.3.1 Invoked Computing: *Zerroug et al., 2011*

Invoked Computing is a current work in progress developed by Zerroug et al. [2011]. The vision behind this concept is to automatically recognize the suggested affordances of a *passive* object and to temporarily transform it into the desired *interactive* object based on multi-modal AR (Augmented Reality). In doing so, an everyday object is enhanced with both audible and visual effects. For this to function, the researchers are working on an in-room AR system which consists of a projector combined with parametric speakers. These speakers create the effect of spatial sound, as if it were coming directly from the object the user is about to transform. Both speakers and projector are mounted on the ceiling beneath a motor so

Invoked Computing turns any passive object into an interactive device through multi-modal augmented reality.

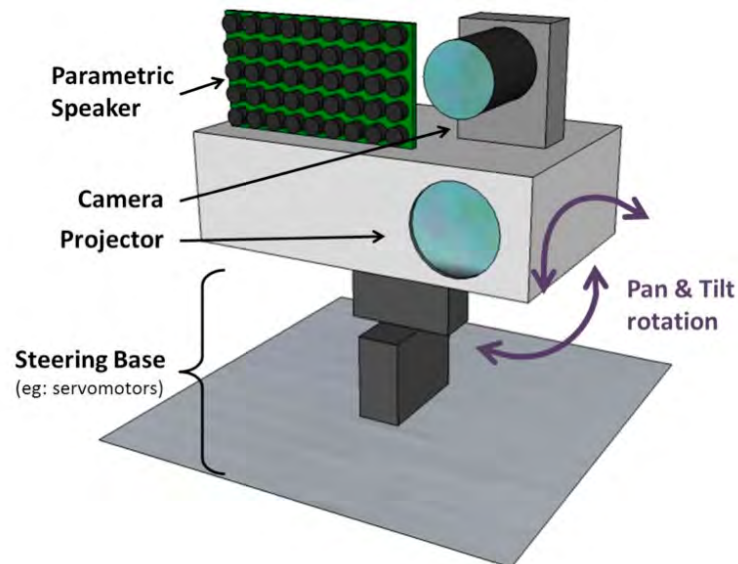


Figure 2.14: The Invoked Computing hardware prototype. A camera, a projector and a parametric speaker are mounted on a steering base which is used for panning and tilt rotation of the prototype. The steering base is mounted on the ceiling of a room. This way, the prototype can augment any object in the room.

that any object in the room can be turned instantly into an interactive device. Figure 2.14 shows the conceptual setup.

Zerroug et al. developed a software kit that can transform a pizza box into a laptop and a banana into a phone.

Although Invoked Computing is still a concept to be realized, the research team developed an early software framework that is capable of automatically transforming a restricted set of objects. As an example, the system can identify a pizza box. When it is opened, a virtual keyboard is projected onto the bottom of the box. Besides, a virtual screen is displayed on the inner side of the lid, so that the box imitates a laptop computer. A second example is a banana that is transformed into a handset through parametric speakers when a user holds the fruit close to mouth and ear.

The researchers are planning to implement an intelligent affordance detection system.

In a future implementation, the research team plans to integrate a completely automatic and intelligent affordance detection system that anticipates how an object is interpreted by the user for appropriation.

Contrasting the concept of Invoked Computing with the work presented in this thesis, both have a different view on the usage of objects. Whereas Co-Optjects target at using everyday objects as haptic controllers for operating any device, Invoked Computing transforms the object itself into a device. Therefore, it not only exploits the haptics of an object, but also adds visual and audible feedback with direct manipulation.

2.4 Related Work at a Glance

Table 2.1 and Table 2.2 present all research projects discussed in this chapter at a glance. They are analyzed regarding prior destination, key feature, type of manipulation, supported types of gesture, exploitation of physical affordances, intensity of object preparation, and technology used.

No.	Project	Destination	Key Feature	Manipulation	Gestures	Affordances	Object Prep.	Technology
1	Mouseless	instant control	eliminates the dependency on a physical computer mouse	indirect	☞, □	–	○	IR laser, web camera with IR filter
2	Skinput	instant access to information	turns the human skin into an interactive surface	direct	☞	–	–	piezo sensors, sound processor
3	WUW	instant access to information	enables mobile access to information displayed on objects	direct	☞, □	–	○	pico projector, pocket web camera
4	Cloth Displays	instant access to information	turns objects into interfaces by adopting their shape	direct	☞, ↔↔↔	○	–	projector, object tracking system based on retro-reflective markers
5	OnObject	object tagging	links acceleration-based gestures to objects and adds audio playback	direct	☞	○	○	RFID tag reader with accelerometer, microphone, and speaker

Table 2.1: Related research projects at a glance (1/2).

Legend: – not/little, ○ medium, + strong, ☞ touch, □ non-stationary 2D movement, ■ stationary and non-stationary 2D movement, □ 3D movement, ➡ acceleration, ↔↔↔ deformation


















No.	Project	Destination	Key Feature	Manipulation	Gestures	Affordances	Object Prep.	Technology
6	Display Objects	rapid prototyping	adds visual style and virtual interfaces to passive mockups	direct	 , 	o	+	projector, Vicon tracking system
7	SketchSpace	rapid prototyping	simulates physical sensors that can be quickly added to passive objects	direct	 ,  , 	+	-	projector, depth camera
8	iCon	instant control	enables control of desktop applications with everyday objects	indirect	 , 	+	o	web camera, paper-based fiducials
9	Invoked Computing	instant control	<i>concept: automatically detects physical affordances of an object for making it interactive</i>	direct		+	-	multi-modal AR
10	Co-Objects	instant control	turns everyday objects into physical controllers for remote target systems	indirect	 ,  , 	+	o	depth camera, web camera, paper-based fiducials

Table 2.2: Related research projects at a glance (2/2).

Legend: – not/little, o medium, + strong,  touch,  non-stationary 2D movement,  stationary and non-stationary 2D movement,  3D movement,  acceleration,  deformation

Chapter 3

Diary Study: A Primer To Everyday Object Interaction

Investigating appropriation of everyday objects first requires to identify which objects are within a person's reach. As the term "everyday objects" suggests, we can expect a common set of objects to be in existence that people come in touch with in quotidian life. This common set may vary depending on culture, gender, and age. Take a lipstick, for example. For an adult woman of western culture, it is likely to be an object used daily. Yet, for women of eastern culture, daily usage of a lipstick could be less usual and concerning men, it is not used at all.

Everyday objects are artifacts people come in contact with in daily life.

In order to define this common set of objects that surround us in daily life, we conducted a diary study based on the recommendations by Lazar et al. [2009] and Preece et al. [2004]. The true power of this type of study is that it can collect data from each user over a period of time without giving the user the feeling of being looked over the shoulder. Besides, data can be collected at different locations and at different times – thus ideal prerequisites for investigating objects of daily usage.

A diary study collects data throughout the day.

3.1 Procedure

People were asked to photograph objects within their reach throughout the day.

The diary study was conducted for one week in May, 2011. Participants were asked to take pictures at least six times per day of objects currently within their reach, thus ensuring, that they were captured independent of a particular location or time. It is important to remark that we did not only ask the users to take pictures of objects, which they were currently using. As mentioned in the introduction of this thesis, IUIs encompass the ubiquity of objects. This does not necessarily mean that an object must be frequently used. Take a family portrait, for instance. Put in a photo frame, it stands on a person's desk, without being actually grasped every day. Yet, this object may be interesting to consider for appropriation as it can be assigned a new role.

Users were asked to take pictures while being at work and while spending leisure time.

As the objects around someone at work may be different from objects within reach after work, we asked users to take pictures on both a workday and on a work-free day, e.g., during the weekend. We offered an optional SMS reminder service that helped to remind people of taking pictures at several times throughout the day. Users signed up online for determining on which days to be reminded and for collecting some demographic data such as age, gender, and profession. The users were not limited in the amount of pictures to be taken. Submission was done by e-mail.

Objects on the pictures were identified, categorized, and finally ranked regarding their occurrence.

For the analysis, the pictures were first stitched into related scenes. For each scene, the shapes of all objects that were clearly visible and identifiable were roughly traced (cf. Figure 3.1). For each user, a "qualitative counting" was done for both scenarios, i.e., workday and leisure time pictures: If a certain object, e.g., a bottle, appeared at least once, it was counted once for all occurrences. We did not count each occurrence as this was hard to do for some objects. Take a stack of paper sheets or towels, for instance. Depending on the image resolution it was nearly impossible to recognize each instance of that object. Having finished qualitative counting, the objects were attributed categories for clarified summarization. For example, different kinds of paper objects, such as calendars, paper

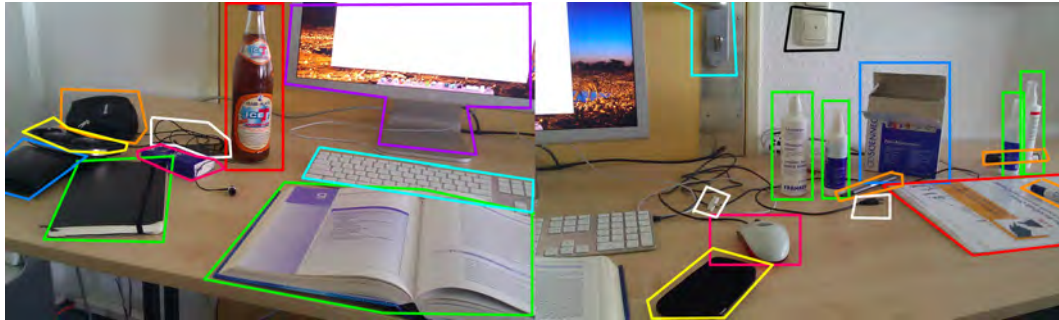


Figure 3.1: Sample picture taken for the diary study. Two pictures have been stitched into a scene and the shapes of all identifiable object have been traced. This picture was taken in a work environment.

sheets, and thin cardboard items were classified as paper. Finally, we created a ranked summary regarding object occurrences assigned to each category for both scenarios.

3.2 Participants

A total of 19 users took part in the diary study (aged 20-62, $M = 29.26$, $SD = 8.93$, two females). The users had various professional backgrounds represented by a major, an architect, a student of arts, project managers, a student of business administration, people involved in media, and one person from the automobile industry. Yet, 42% of all candidates were computer science students or research assistants. Seven people used the SMS reminder service.

Participants had various professional backgrounds.

Although the user group does only represent people from western culture, we have at least people with various professional backgrounds which therefore contributes to a good variety as regards work environments.

3.3 Results

The participants sent in 360 pictures in total. Collectively, 15 candidates handed in photos for both workday and

A total of 360 pictures were collected.

	<p>work-free day, three participants took pictures only representing leisure environments, and one user only sent in images captured on a workday.</p>
Altogether, 497 objects were assigned to 98 categories.	<p>Altogether, 497 objects were counted which were attributed to 98 distinct categories, from which 67 came from the pictures taken at work, and 88 contributed to leisure environment. In total, 57 categories could be found in both scenarios. Hazardous objects, such as manual shavers, were excluded from our analysis, as these objects should not be appropriated, since they can cause harm in case they are not used as intended. Anyway, such objects were not photographed frequently.</p>
Participants took close-up pictures and wide shots.	<p>Pictures were taken differently. Most candidates photographed the entire environment they were in while taking the picture. Hence, these pictures displayed many objects at once. No-one used a panorama shot. Instead, multiple pictures were taken at different view points for a room or environment. As such a series of photos had a lot of overlapping parts, they were easy to stitch together to related scenes. Others used single-shot exposure for taking detailed close-ups of each object individually.</p>
A ranked summary for object occurrences was created.	<p>A tabular summary of the objects counted most frequently is given in Table 3.1, separated by work and leisure environment. The table shows that different groups of objects surround people in quotidian life. Basically, these are furniture, dishes, paper-based objects, things we carry in our pockets, and tools for the office. As regards furniture, tables were observed most frequently. However, interaction with tables is already covered by multi-touch research [Han, 2005]. Therefore, we will not consider furniture in our work.</p>
Dedicated input devices were also identified among frequently observed objects.	<p>Dedicated input devices, such as keyboards, mice, and remote controls are also at people's reach. One might question whether there is actually a need for investigating IUIs. As regards mouse and keyboard, this might be true for office environments and office-related tasks, since these devices are frequently available. Similarly, remote controls are often available in leisure environments. Yet, it is interesting to find out whether other objects with different affordances compared to these push-dominant controllers</p>

Object Occurrences				
Rank	Category	Leisure	Work	Σ
1	Table	16	12	28
2	Box/Container	15	10	25
3	Paper	10	13	23
4	Bottle	12	10	22
5	Drinking Vessel	11	9	20
6	Pen	8	10	18
6	Underlay	11	7	18
7	Cable	7	10	17
8	Book	7	9	16
9	Laptop/Computer <small>incl. monitor</small>	5	10	15
9	Phone	9	6	15
10	Cloth	12	2	14
10	Cutlery	11	3	14
11	Clothes	10	3	13
12	Plate	11	1	12
13	Chair	5	6	11
13	PC Keyboard	2	9	11
14	Bag	6	4	10
14	Lamp	7	3	10
14	Wallet	7	3	10
15	Key	8	1	9
15	PC Mouse	2	7	9
15	Spray Can	7	2	9
16	Cushion	7	1	8
16	Door	6	2	8
16	Puncher	2	6	8
16	Remote Control	8	0	8
16	Stapler	0	8	8
17	Watch	6	1	7
18	Stick	0	6	6
19	White Board	0	5	5

Table 3.1: Object occurrences identified for the diary study. This table shows the number of object category occurrences in the given environment (work vs. leisure). Whenever an object belonging to a category was observed at least once for a participant, the corresponding leisure/work counter was incremented ($N = 19$). The table is ranked by the overall occurrences and shows only items that occurred overall at least five times based on qualitative counting.

could not do a similarly good job concerning the manipulation of an interface. A bottle, for example, has a completely different physical shape and comes with a turnable cap instead of push buttons. Could screwing a cap off a bottle, for instance, be a more comfortable way of controlling a dimmable light than operating a switch? Although some dedicated devices have become omnipresent to us, it is still worth finding out how everyday objects perform as their temporary substitutes for situations like when the dedicated device is broken, temporarily unavailable, or annoying to carry around as motivated by Mistry and Maes [2011].

Carried items were photographed more often in leisure environments than in work environments.

Objects captured in the work environment were mostly desktop items that were typical for office jobs. Objects photographed in leisure environments were typically found in the living room, bathroom, or kitchen. In addition, among these objects, seven items were things people usually carry with them: clothes, phones, pens, keys, wallets, bags, and watches. Interestingly, people tended to photograph carried items more often while being in their leisure environment (up to seven items) compared to the working environment (up to three items) although these objects are completely independent of the environment. Our guess is that people were more busy during work times and that they felt that their personal (and mobile) objects do not belong to the work environment, thus they only photographed them at home. The low number of photographed phones was probably due to the fact that 16 out of 19 people used a mobile phone with an included camera for taking the pictures which consequently cannot appear on the photos.

The data from the diary study is used for follow-up studies.

Using this statistical data as primer for the follow-up study, the next goal was to see which of these frequently observed objects were actually adequate for being appropriated and how people *would* manipulate them given the scenario of an intelligent room being able to interpret any kind of object manipulation for appropriation.

Chapter 4

Wizard of Oz Study: Interaction Patterns for Instant User Interfaces

In the preceding study, we identified the most frequent objects available within a person's reach. Given a set of IUI-conform scenarios, the next step is to analyze (1) *which* of these objects are considered appropriate for appropriation and (2) *how* users actually manipulate them. Therefore, we conducted a Wizard of Oz (WOZ) study, giving participants full freedom in appropriating objects of their choice for controlling remote target systems presented in small scenarios. Overall, we are looking for object-gesture-action mappings created by the testers: Given a certain scenario in which the dedicated input device of a remote target system is missing, which *gestures* do the users apply on one or more objects within reach for accomplishing a prescribed goal dependent on a certain action? Making use of WOZ prototyping, our goal was to gain insight into the users' mental models on object appropriation in the sense of IUIs by analyzing general interaction patterns.

For finding out how people appropriate everyday objects in terms of IUIs, we conducted a Wizard of Oz study.



Figure 4.1: Object variants for a drinking vessel. A drinking glass (left) is rigid, a paper mug (middle) is lightweight, flexible, and has a removable lid. A mug (right) has similar rigidity as a glass but additionally features a handle.

4.1 Setup

Due to variation in physical affordances, we included different object variants in the study.

The study was conducted in a spacious room equipped with large window sills and tables arranged in the shape of an L. The storage surface served as exhibition space for presenting a large set of everyday objects based on the top ranking from the diary study (cf. 3—“Diary Study: A Primer To Everyday Object Interaction”). As physical affordances played an important role in this study, some objects existed in more than one version, varying in their material. For a drinking vessel, for instance, we included a drinking glass, a porcelain cup and a paper mug (cf. Figure 4.1). All items differ in weight, elasticity and surface texture.

The user had an unobstructed view on all objects available for appropriation.

In front of the exhibition space, we put a table with a monitor serving as the user’s workspace. From this perspective, the participant had an unobstructed view of all objects and the display. The monitor served as information display for the user illustrating the scenarios and the demanded tasks. Space between table and exhibition space was empty, enabling the user to easily access all objects.

The user’s workspace was captured by camera.

For the purpose of qualitative analysis, we placed a digital video camera at both sides of the user’s workspace. The user’s manipulation on the everyday objects alongside

with the participant's spoken thoughts were recorded from two perspectives. The complete hardware setup is depicted in Figure 4.2.

4.2 Procedure

At the beginning of the experiment, the user was given the chance to inspect all objects from the exhibition space for being aware of which objects were available and by which physical properties they were characterized. For each tester, the arrangement of the object in the exhibition space was randomized.

Before the study started, each user was given the opportunity to inspect all objects.

At first, the user was informed to encounter short virtual scenarios dealing with interactive systems to be controlled with everyday objects (free of choice) instead of using a dedicated input device. Furthermore, the participant was introduced to the WOZ ideology: The user could expect the study room to be intelligent enough to recognize any (physical) gesture applied on objects. It would also be able to share the user's interpretation, i.e., whenever the user applied a certain gesture on an object for solving a task, the system would be able to recognize the gesture and execute the current task.

The user could expect the study room to be intelligent enough to recognize any physical object manipulation.

The scenarios were split into two blocks; one representing small GUIs, each representing a composition of basic UI widgets, the other one illustrating interaction with usual remote target systems, i.e., holding a presentation, watching television, playing a video game, operating the light, and a futuristic scenario called "interaction with a virtual shop window". The sequence of scenarios within a block was randomized, as was the order of the two blocks counter-balanced. Each scenario consisted of a sequence of tasks that were both read aloud by the principal investigator and printed on the monitor. For better imagination of the virtual scenarios, digital sketches were presented alongside with the task descriptions. The sketches were created with a [Wacom Cintiq interactive pen display](http://www.wacom.com/en/Products/Cintiq.aspx)¹; the

The scenarios were divided into two blocks and the tasks were visually augmented with sketches.

¹www.wacom.com/en/Products/Cintiq.aspx



Figure 4.2: Study setup for the Wizard of Oz study. From her workspace (table, foreground), the user had an unobstructed view on all objects (background). Object manipulation for solving the tasks (displayed on the monitor) was captured by two cameras to the left (beneath the monitor) and right).

presentation of sketches and tasks was done with slides using [Apple Keynote](https://www.apple.com/iwork/keynote)².

After a task was accomplished, the sketch was updated for providing visual feedback.

After a task was accomplished, the sketch was updated to indicate the new system state. Figure 4.3 shows an example for three steps of incremental visual feedback for the television scenario. A complete set of all sketches can be found in A—“Appendix for the Wizard of Oz Study”. Although the ambient room was intelligent and thus able to interpret the user’s gestures, real time feedback was not supported. Live feedback was available in the final study of this work (cf. 5—“Testing Everyday Object Interaction with an Interactive Prototype”).

Subsequent tasks were not perceptible beforehand.

Within a scenario, tasks were not randomized as a subsequent task was dependent on the previous one. Moreover, the tasks were faded-in one after another so that the

²www.apple.com/iwork/keynote

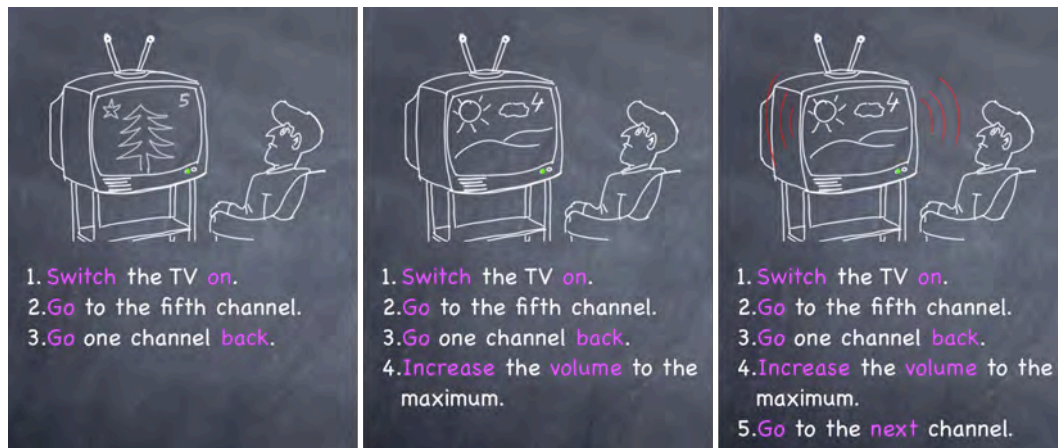


Figure 4.3: Incremental visual task feedback for the TV scenario. Only *after* a task was accomplished, the sketch was updated (sequence: from left to right).

participant was unable to see which tasks were to follow. This was done for simulating spontaneity, illustrated by the following example:

GOALS ARISING SPONTANEOUSLY WHILST HOLDING A PRESENTATION:

A presenter forgot his remote control. During the presentation, he wants to go to the next slide (goal I). He appropriates a pen that shows the next slide by pushing the pen button. Yet, after some time, he might need to recall a previous slide (goal II), so that he feels the need for a second button, which the pen does not offer. If he had thought of this before, he probably would have chosen a different object at hand. However, in spontaneous situations, one does merely focus on the current goal (I) and not at possible future goals (II). For simulating this behavior, the tasks presented in the WOZ study were faded-in step by step.

Example:
Goals Arising
Spontaneously whilst
Holding a
Presentation

Having explained the basic study procedure, the first block of scenarios was presented to the participant. For accomplishing a presented task, the user was asked to bring one or more³ objects from the exhibition space to the workspace and execute the gesture(s) coming to her mind

The users were free to pick objects, and they were free to put them back at any time.

³The user was free to decide on the number of objects needed.

that would solve the task. At any time, the user was free to put objects back and fetch new ones.

Next to video recordings, qualitative user feedback was collected with the aid of a questionnaire.

After a scenario had been completed, all objects were put back to the exhibition space for neutralizing the setting. The qualitative observations were topped of with a final questionnaire, basically consisting of questions conform to a five-point Likert scale, dealing with task performance, choice of objects, and general impression on IUIs. A score of 1.0 means “strongly disagree”, whereas a value of 5.0 represents “strongly agree”.

4.2.1 Objects

Based on the findings of the diary study, the set of manipulatable objects was adapted to the environment where a scenario is likely to be encountered.

Referring to the results of the diary study, the set of choosable objects was adapted to the environment in which a scenario is likely to be encountered. For example, activities like watching television, playing a video game, etc., are mostly experienced outside the office, whereas interaction with GUIs mostly takes place at work/in the office. Thus, objects that were counted in leisure environments were available for remote target system scenarios, whereas objects counted on the pictures taken in work environments were available for GUI scenarios. As talks are often given in meeting rooms or at conferences in which typical office equipment, such as a stapler, a puncher, etc., are not available, we attributed this scenario to leisure environment, matching the corresponding object set better than for work environments. Besides, the object set for the virtual shop window scenario was reduced drastically to objects people usually carry in their pockets, as this scenario takes place outdoor.

In total, 63 items were available in this study.

In total, 63 items from 30 different object categories were at disposal in this study. A full list of all objects that were available for each scenario is given in Table 4.1.

Scenarios			Object Variants
A1–A4	A5	B1–B7	
Bag	Bag	Bag	<i>clicky-top, twisty-bottom mobile (clamshell), stationary</i>
Pen	Pen	Pen	
Phone	Phone	Phone	
Pocket Tissues	Pocket Tissues	Pocket Tissues	
Clothes	Clothes		
Key	Key		
Wallet	Wallet		
Watch	Watch		<i>shoes, jeans, shirt, pullover</i>
Book		Book	
Bottle		Bottle	
Box/Container		Box/Container	
Cable		Cable	
Drinking Vessel		Drinking Vessel	
Paper		Paper	<i>softcover, hardcover</i> $\{small, big\} \times \{glass, plastic\}$ <i>cylindric</i> <small>(rigid)</small> , <i>cuboid</i> <small>(flexible)</small>
Underlay		Underlay	
Cloth			
Cushion			
Cutlery			
Desk Lamp			
Plate			
Remote Control			<i>porcelain cup, drinking glass, paper mug</i> <i>mouse pad, desk pad</i> $\{metal, plastic\} \times$ $\{fork, knife, spoon\}$ <i>metal, plastic</i> <i>glass/plastic flacon, spray can</i> <i>three-button, scroll wheel</i>
Spray Can			
		Keyboard	
		Mouse	
		Glue Stick	
		Laptop	
		Monitor	
		Puncher	<i>incl. three magnets and three colored marker pens</i>
		Stapler	
		White Board	

Table 4.1: Available object types per scenario. Objects for scenarios A1–A4 were based on the most frequently observed items available in leisure environments according to the diary study; objects for scenario A5 represented wearable items and were a subset of objects used in A1–A4. Objects for scenarios B1–B7 were based on items at hand in work environments. The remote control was not available in scenarios A1 and A2. PC mouse and PC keyboard were only allowed to be used as general multi-button devices.

4.2.2 Scenarios

For the reader's comfort, a detailed description of each scenario is given step by step in the evaluation part of this study. Before doing a profound per-scenario analysis, we substantiate why we believe these scenarios to be representative for IUI research.

Block A: Remote Target Systems

Remote target systems, as presented in the scenarios, are usually dependent on a dedicated input device.

Scenarios from block A, i.e., holding a presentation, watching television, playing a video game, operating the light, and interacting with a virtual shop window, have been tested in this study because we believe that they are typical situations that benefit from the IUI metaphor. All of them are dependent on a dedicated input device, such as a remote control, game pad, or switch, that can happen to be broken, out of reach, or generally unavailable. Although the virtual shop window scenario is of futuristic nature, it is still valuable for this study as it represents an outdoor scenario and tests three-dimensional navigation – unlike all other scenarios.

Block B: UI Widgets in GUIs

GUI scenarios were used to test basic control of UI widgets including focus switching with the aid of appropriated objects.

Regarding scenarios from block B, a collection of UI widgets, they do not represent realistic GUIs. Our aim was to test control via appropriation for each basic type of UI widget, i.e., how people manipulate a widget using everyday objects plus how they do switch focus between multiple instances. Apart from these “simple” GUIs, we included a spreadsheet interface being of more complex nature. It consisted of multiple cells and a horizontal and a vertical scroll bar. Again, this GUI was not a representative one and was only used for intensively testing focus switching between different UI elements. Hence, these scenarios were tested for being able to extract basic appropriation patterns for distinct UI widgets. Appropriation for a real GUI, such as Apple QuickTime Player that combines multiple wid-

gets of different kinds, was tested in our final study (cf. 5—“Testing Everyday Object Interaction with an Interactive Prototype”).

We believe that the WOZ study supported investigation of everyday object appropriation in the sense of IUIs for the following reasons:

We considered the WOZ approach supportive.

- Using a combination of camera observation and think aloud, the study helped to gain insight into the users’ mental models towards IUIs, i.e., it helped to understand how people bridge the Gulf of Execution using everyday objects when the dedicated tool is missing.
- The study supported the aspect of spontaneity, as it presented tasks step by step which was forcing the user to reconsider the choice of object as this would also be necessary in a real scenario, i.e., outside the study room.
- Both blocks of scenarios represented settings in which a dedicated input device is usually needed. Although B-scenarios dealt with artificial GUIs, testing appropriation for basic UI widgets could allow for predicting everyday object interaction as regards GUIs composed of these widgets. In addition, it helped to gain insight into focus switching methodologies.
- Finally, based on the results of the previous study, this user test represented a realistic choice of objects within reach, depending on the scenario.

4.3 Participants

A total of 18 participants, aged 20–28 ($M = 24.39$, $SD = 2.15$), two females, participated in this study. Besides a teacher, a medicine student, and a student of business administration, only computer science students took part in the experiment.

In total, 18 participants took part in the study.

4.4 Evaluation

4.4.1 Method of Analysis

Whenever an object-gesture mapping was observed for a task at least three times, we identified it as a pattern.

Subsequently, we present the results of our qualitative analysis. For each scenario, we identified the most frequent object-gesture-task mappings, also referred to as patterns: Whenever the same task solution was observed in three or more users, it was denoted by a pattern. Furthermore, we interpreted the role attributed to an object, e.g., when a user tried to identically replace a remote control.

For reasons of spontaneity, only the first solution presented by a user was regarded.

For final evaluation, we only counted the first complete solution for a scenario. Although some users presented alternative solutions, we were only interested in the most spontaneous pattern, as the word “instant” in IUIs suggests. In time-critical situations, such as illustrated in the previously presented example (cf. 4.2—“Procedure”), spontaneity is an important factor in case the dedicated input device is missing and when a quick workaround is needed to bridge the arising the Gulf of Execution. In this context, physical affordances contribute to finding a solution: They suggest the most natural – and therefore the *first* – way of how the object can be manipulated. Therefore, secondary solutions were not of interest.

Incomplete solutions were filtered out in case the user also presented a complete set of mappings for the scenario.

In addition, incomplete and invalid task solutions were filtered out. An invalid solution would be, for instance, making use of speech recognition as our (fictional) ambient system was only capable of recognizing and interpreting (physical) object manipulation. Some results from the feedback form may be anticipated, where appropriate. Questions are referenced by Qn , where n is the number of the question as listed in the questionnaire. Detailed analysis on the questionnaire, however, is listed in 4.4.5—“Qualitative User Feedback”. The complete form is printed in A—“Appendix for the Wizard of Oz Study”.

We implemented a software for synchronous video playback.

For efficient analysis of the video recordings, we implemented a small software, called “DualPlayer” that assured synchronous playback and pause of the two video streams recorded per person.

4.4.2 Scenario Analysis

For evaluation of the WOZ prototype, we analyze each scenario. Besides, subjective feedback from the testers is analyzed. Finally, we draw a conclusion on these findings and generalize them.

Most users were able to find object-gesture-action mappings right from the beginning for each scenario, i.e., they did not need to try out several artifacts before finally committing oneself to a specific object. Hence, appropriation was spontaneous. The order of blocks (A,B vs. B,A) did not have any impact on the user as no abnormality in object-gesture usage could be observed. If not otherwise indicated, the objects used by the participants were put on the table (i.e., the user's workspace) for manipulation.

Most users were able to find task solutions without prior experimentation.

4.4.3 Scenarios A: Observations

A1: Holding a Presentation

The user is about to present a set of digital slides. However, she forgot to bring a remote control and therefore is unable to navigate through the slide set.

Precondition: The presentation has been started.

Tasks and Patterns

1. *Go to the next slide.*
 - (a) **ball pen:** push the pen button once
 - (b) **mobile phone:** push the right button on the directional pad
 - (c) **hardcover book:** flip one page forward

2. *Go to the next slide.*
 - (a) **ball pen:** push the pen button once
 - (b) **mobile phone:** push the right button on the directional pad
 - (c) **hardcover book:** flip one page forward
3. *Go to the previous slide.*
 - (a) **ball pen:** double-push the pen button
 - (b) **mobile phone:** push the left button on the directional pad
 - (c) **hardcover book:** flip one page backward
4. *Skip five slides forward.*
 - (a) **ball pen:** push the pen button five times
 - (b) **mobile phone:** push the right button on the directional pad five times
 - (c) **hardcover book:** flip five pages forward

*

Navigating the slides was push-dominant.

In this scenario, a ball pen or a mobile phone were most frequently used. The participants were trying to substitute the functionality of the missing remote control: Both the pen and the mobile phone are equipped with physical buttons that the testers pushed for flipping the slides. Figure 4.4 (left) shows the mobile phone keypad layout.

Apart from imitating the remote control, two testers appropriated the pages of a book as physical instances of the virtual slides.

Interestingly, two testers rather tried to *physically instantiate* the virtual slides, i.e., the physical object did not substitute a remote control but represented the virtual slides. By browsing the pages of a hardcover book, the users expected the system to flip the slides accordingly. Thus, each physical page represented a virtual slide. This phenomenon follows the principal of natural mapping, as defined by Norman [2002]:

Natural mapping, by which I mean taking advantage of physical analogies and cultural standards, leads to immediate understanding. – Norman [2002]



Figure 4.4: Foldable mobile phone (left) and remote control (right) as available in the study. Both devices are equipped with multiple buttons and feature both a numeric pad and a directional pad.

We substantiate this term – narrowed to the field of IUIs – as *Physical Instantiation*.

PHYSICAL INSTANTIATION (PI):

The act of using a physical object as representative substitute of an either physical or virtual target for remotely directly controlling it. Manipulation of the substitute is supposed to equally effect on the remote original.

Definition:
Physical Instantiation (PI)

Please note that PI only refers to instantiation of the entire target system and not to a remote control. For example, using a mobile phone to replace a remote control for navigating the slides is not an example of PI.

PI refers to a remote target system as a whole.

Moreover, for 16 participants manipulating just one object for accomplishing all tasks was sufficient. The remaining two testers used two objects each. The patterns were applied consistently over all tasks, i.e., users did basically stick to the object chosen at the beginning.

Most testers could handle all tasks with just one object.

Of course, the remote control (cf. Table 4.1) was not available in this scenario.

A2: Watching Television

The user is about to watch TV. Yet, the remote control is out of reach.

Precondition: The TV is in standby mode.

Tasks and Patterns

1. *Switch the television on.*
 - (a) **mobile phone:** flip it open
 - (b) **pocket tissues** (held in portrait format): point at the TV and push the upper right corner of the pack of pocket tissues
 - (c) **mobile phone:** point at the TV and push the center button on the directional pad
 - (d) **ball pen:** push the pen button once
2. *Go to the fifth channel.*
 - (a) **mobile phone:** point at the TV and push button "5"
3. *Go one channel back.*
 - (a) **mobile phone:** point at the TV and push the left button on the directional pad
 - (b) **mobile phone:** push the lower button on the directional pad
 - (c) **mobile phone:** point at the TV and push button "4"
4. *Increase the volume to the maximum.*
 - (a) **mobile phone:** push the upper button on the directional pad multiple times
 - (b) **mobile phone:** point at the TV and hold the right button on the directional pad pushed
 - (c) **bottle:** turn it/its cap clockwise

5. *Go to the next channel.*

- (a) **mobile phone:** point at the TV and push the right button on the directional pad once
- (b) **mobile phone:** point at the TV and push the upper button on the directional pad once

6. *Mute the volume.*

- (a) **mobile phone:** push any (so far unused) button
- (b) **mobile phone:** push the lower button on the directional pad a few times

7. *Switch the television off.*

- (a) **mobile phone:** point at the TV and push the red button (= on/off button)
- (b) **mobile phone:** flip it shut
- (c) **mobile phone:** point at the TV and push the center button on the directional pad.

*

The imaginary television set was most frequently controlled with the aid of a mobile phone. Interestingly, before actually executing a gesture, seven participants regularly pointed with their object at the TV, although this was not necessary, as there was only one “TV” available in the intelligent room. Nevertheless, this proved that people were strongly trying to find an almost identical replacement for the original remote control. Indeed, no-one applied PI.

Participants used the mobile phone as remote control replacement and pointed at the TV.

At the beginning, also pocket tissues and ball pens were used for turning the TV on. With the second task, however, these objects were hardly used, as the mobile phone dominated and worked better for the intermediary tasks due to its richness of buttons. Hence, object usage was not consistent over the tasks. The shape of the phone, its buttons and the pointing proved that the users were looking for an object similar to a TV remote control.

At the beginning, a ball pen or pocket tissues seemed sufficient, but these objects were unsuitable for subsequent complex tasks.

As regards volume control, four users did not use buttons for this task. They were looking for round-shaped

Changing the volume by turning the bottle (cap) was identified as a pattern.	objects that could be turned, such as the cap of a bottle for continuously regulating the volume. The distribution regarding usage of buttons vs. turning for volume control matched the results of the questionnaire, as 14 testers said that they perceived changing the volume either discrete input, i.e., countable individual levels, or they did not care about whether using a discrete or continuous control (Q8).
Only two participants used multiple objects simultaneously.	In this scenario, eleven people just needed one object to control all tasks. Only two participants used multiple objects simultaneously for switching the channels. For example, one participant placed a fork on the table with tines showing up, while pressing the button of a ball pen for going to the next channel. Letting the tines face the table surface changed the channel direction: pushing the pen button would result in going one channel back.
Overall, simultaneous usage of multiple objects was rarely observed in this study.	Yet, for all A- and B-scenarios, simultaneous usage of two or more objects was rarely observed. Most participants preferred using one object at a time at maximum. Trivially, the TV remote control was not available in this scenario.

A3: Playing a Video Game

The user is about to play a 2D jump and run game with a friend. However, she has just one controller at hand.

Precondition: See Figure 4.5.

Tasks and Patterns

1. *Jump and move to the right to land on the tube.*
 - (a) **remote control** (held in landscape format): push a button on the left side (jump) and a rightmost button on the right side (move right)
 - (b) **remote control** (held in portrait format): push the upper button (jump) and push the right button (move right) on the directional pad



Figure 4.5: Precondition for scenario A3: *Playing a Video Game*. The user was asked to navigate Super Mario (avatar) through the scene.

- (c) **spray can:** push the button on top of the can (jump) and move the can to the right (move right)

2. *Jump and fire the enemy.*

- (a) **remote control** (held in landscape format): push the upper button on the directional pad⁴ (jump) and press a rightmost button on the right side (fire)
- (b) **remote control** (held in portrait format): push the upper button on the directional pad (jump) and any (so far unused) other button, e.g., “red”, which was identified with the color of fire
- (c) **mobile phone** (held in landscape format, flipped open): move the thumb on the display up and push a button on the numeric keypad

⁴corresponds to the right button on the directional pad when the device is held in portrait mode

3. *Turn left and fall down to stand next to the tube.*
 - (a) **remote control** (held in landscape format): push any leftmost button on the left side
 - (b) **remote control** (held in portrait format): push the left button on the directional pad
 - (c) **mobile phone** (held in landscape format, flipped open): move the thumb over the display to the left
4. *Run directly to the very left of the screen.*
 - (a) **remote control** (held in landscape format): hold any leftmost button on the left side pushed
 - (b) **remote control** (held in portrait format): hold the left button on the directional pad pushed
 - (c) **mobile phone** (held in landscape format, flipped open): swipe the thumb over the display to the left and hold the finger in that position
5. *Run directly back to the tube and stop.*
 - (a) **remote control** (held in landscape format): hold any button on the right side pushed
 - (b) **remote control** (held in portrait format): hold the upper button on the directional pad pushed
 - (c) **mobile phone** (held in landscape format, flipped open): swipe the thumb over the display to the right and hold the finger in that position

*

The testers tried to imitate a game pad by holding button-equipped devices in landscape format.

Like the preceding scenarios, playing a two-dimensional jump and run game was strongly push-dominant as the testers generally tried to replace the missing game controller with everyday objects: The remote control and the mobile phone were most frequently chosen and held in landscape format. Typical game pads, such as known from [Sony PlayStation](http://www.sony.com/playstation)⁵ or [Microsoft Xbox](http://www.xbox.com)⁶ resemble that shape.

⁵www.sony.com/playstation

⁶www.xbox.com

The patterns used in this scenario were most inconsistent. For example, some users switched the button used for letting the avatar jump from Task 1 to Task 2. Such inconsistencies were not realized by the testers and were probably due to missing live feedback.

Due to missing live feedback, users were unaware of using different buttons for the same action.

Interestingly, the testers did not assign just one button to a certain action, but rather generalized a group of buttons of close proximity and linked it to an action (cf. Figure 4.4, right). Only for the case the directional pad was used, the participants used exactly one button for movement, firing, or the like. The mobile phone was held in landscape format and therefore – like the remote control – showed that the testers were trying to imitate a game pad, which are usually designed to be held in the same way. For movement of the avatar, the mobile phone users exploited the smooth display as touchpad: The avatar was meant to move likewise to the thumb moved on the small phone screen (cf. Figure 4.4, left).

Some participants appropriated the display of the phone as touch-sensitive display.

Interestingly, there was a small tendency for applying PI in this scenario: Five testers used everyday artifacts for either some or all tasks as referential objects. One participant, for instance, used a pair of shoes and moved them to the left or right to control the avatar.

PI was also observed.

Besides, more than simple tactile feedback was looked for: One user held a bottle vertically with two hands to control the avatar. Tilting the bottle to the left caused the water to flow to the same direction and gave the user direct feedback by increasing weight. In this scenario, two objects were used at maximum per person; 13 testers just needed one object for all tasks.

Two objects were used at maximum per person for this scenario.

Interestingly, although the game was a two-dimensional sample, those who applied PI automatically transferred the setting into a 3D environment: For instance, one user tried to rebuild the scene with objects by putting a squared box on the table that symbolized the tube depicted in the screenshot. A spray flacon represented the avatar. For jumping on the tube, the participant simply put the flacon on the box (cf. Figure 4.6). As three-dimensional movement was perceived continuous input by most people (15 vs. three participants, who considered it being either

One user applied PI by rebuilding the scene with a box and a flacon.

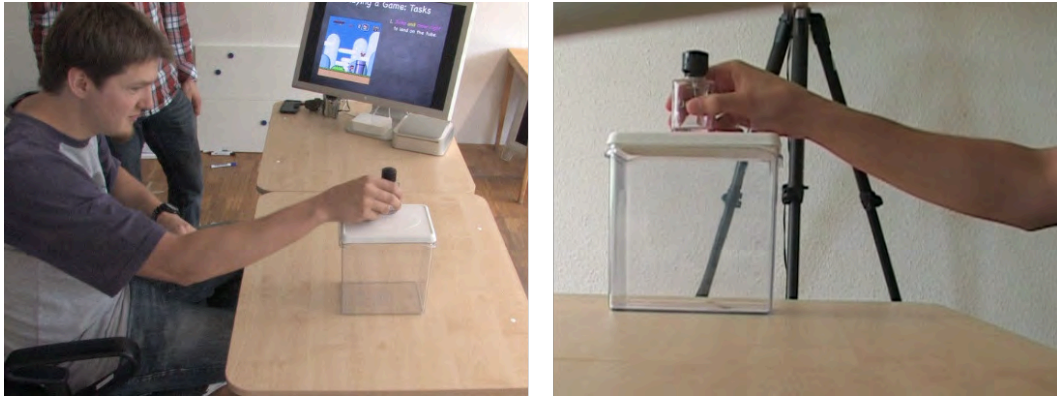


Figure 4.6: Applied Physical Instantiation for scenario A3: *Playing a Video Game*. The user appropriated a spray flacon as avatar, whereas the tube depicted in the screenshot was represented by a box.

discrete or did not care, cf. Q5) compared to 2D navigation (eleven testers considered it continuous input; the others considered it discrete or did not care).

We claim, that PI is likely to be applied when continuous input is involved.

The proportion of slightly more applied PI in this scenario matches our interpretation that this paradigm is more often observed when continuous input (such as three-dimensional movement) is involved.

A4: Operating the Light

The user is about to control two light sources. Yet, the light switches are out of reach.

Precondition: Both light sources are switched off.

Tasks and Patterns

1. *Switch light source B on.*

- (a) **ball pen:** point at source B and push the pen button once

- (b) **remote control**: point at source B and push the power button once
 - (c) **mobile phone**: push button “2”
2. *Illuminate light source B to maximum brightness.*
- (a) **drinking vessel/bottle with cap**: turn the cap clockwise
 - (b) **ball pen**: unscrew the pen head from the pen body
3. *Dim light source B darker.*
- (a) **drinking vessel/bottle with cap**: turn the cap counter-clockwise
 - (b) **ball pen**: twine the pen head and the pen body together
4. *Switch light source A on.*
- (a) **ball pen**: point at source A and push the pen button once
 - (b) **mobile phone**: push button “1”
5. *Switch light source B off.*
- (a) **ball pen**: point at source B and push the pen button once
 - (b) **mobile phone**: push button “2”

*

The light scenario was also push dominant, accompanied by a strong use of point- and turn-gestures. The lights were dimmed by turning an entire bottle or its cap. Pushing was mostly done for switching the lights on/off with a ball pen button, a remote control, or a mobile phone. Pointing for selecting a light source was done with pens and bottles. Alternatively, users assigned a button for activating a source. Those who assigned buttons “1” and “2” of the keypad may have been influenced by the task formulations which identified the light sources with letters A and B. The remote control was less influential for this setting.

Pointing was used for switching focus from one light source to another.

	<p>This scenario involved the highest object usage per person ($M = 1.61$ objects per user, $SD = 0.78$).</p>
PI was applied once.	<p>Only one person partially applied PI by picking a desk lamp: The small instance was to represent light source A or B, respectively. All other participants used their chosen objects as remote controls or remote light switches, respectively.</p>
Compared to A1–A3, this scenario was perceived least easy.	<p>Overall, finding objects and according gestures was perceived least easy comparing this scenario to A1–A3 (Q6, Q7, cf. Figure 4.10). Continuously dimming paired with switching the focus between different light sources was considered challenging.</p>

A5: Interacting with a Virtual Shop Window

The user is in a city and faces a large wall display showing a 3D-rendered pair of glasses. The user is about to rotate the virtual object from at distance for inspecting it.

Precondition: The glasses are facing the user.

Tasks and Patterns

1. Turn the glasses to the backside.

- (a) **mobile phone** (held in landscape format): turn it to the backside
- (b) **ball pen**: point at the virtual glasses and move the pen to the left or to the right
- (c) **pocket tissues** (held in landscape format): turn the pack to the backside
- (d) **wallet** (held in landscape format): turn it to the backside

2. *Flip the sides of the glasses open.*
 - (a) **wallet** (held in landscape format): flip it open
 - (b) **mobile phone** (held in landscape format): flip it open
 - (c) **watch**: flip the watch strap open
3. *Turn the glasses 90° to the left.*
 - (a) **mobile phone** (held in landscape format, flipped open): turn it 90° to the left
 - (b) **wallet** (held in landscape format, flipped open): turn it 90° to the left
 - (c) **ball pen**: point at the virtual glasses and move pen to the left or to the right
4. *Tilt the glasses 45° up.*
 - (a) **mobile phone** (flipped open): tilt it 45° up
 - (b) **wallet** (flipped open): tilt it 45° up
 - (c) **ball pen** (held horizontally): tilt it 45° up

*

In contrast to the previous scenarios, the virtual shop window setting was dominated by applied PI: In total, 17 participants titled, rotated, and flipped either a mobile phone or a wallet intending the virtual counterpart to move 1:1. As regards video analysis, causing the sides of the virtual glasses to be flipped open was perceived most challenging. Yet, PI helped best: The participants flipped the mobile phone or the wallet open for controlling the virtual glasses.

in this scenario, 17 users applied PI.

In this scenario, three objects were used at maximum per person; twelve users only needed one item to control the glasses in the virtual shop window.

Most users needed only one item to solve all tasks.

As regards the questionnaire, this scenario was perceived least easy among all A-scenarios (Q6: $M = 3.22$, $SD = 1.17$, Q7: $M = 3.94$, $M = 0.87$ average rating for the ease of finding objects/gestures, cf. Figure 4.10). Indeed,

This scenario was perceived least easy.

Users were looking for an object that resembled a pair of glasses.

Again, PI is more likely to be applied when continuous input is demanded.

one participant gave up in the face of finding an object and a gesture for flipping the sides of the virtual glasses open.

Primarily, people were looking for an identical physical counterpart: a pair of glasses. As these were not available, the testers predominantly picked the wallet or the mobile phone because these objects roughly resembled the basic shape of a pair of glasses: a rectangle. One user said while picking the pocket tissues: *“Since the glasses are principally rectangular, I am looking for objects which have more or less a similar shape.”*. The fact that both the phone and the wallet could be flipped open was of coincidence; people needed some time until they realized that flipping was also applicable to these items. Therefore, the task was perceived difficult. The testers commented that some object properties, such as flippable sides, were too object-specific and were hard to control when using a referential substitute that did not exactly represent this feature.

The observations made in this scenario confirm our previously made assumption: PI is perceived being easier to apply than imitating a remote control in case continuous input is demanded. Indeed, spatial navigation of the glasses was perceived continuous input by all testers (Q8).

4.4.4 Scenarios B: Observations

B1: Buttons

Three buttons (A, B, C) are arranged next to each other.

Precondition: Focus is on button A.

Tasks and Patterns

1. *Push button A.*

- (a) **ball pen:** point at button A on the screen and push the pen button once
- (b) **three-button mouse:** push the left button once

- (c) **keyboard**: hit the enter button
- (d) **magnets** (three magnets put in a horizontal line): touch the leftmost magnet with the forefinger

2. *Push button C.*

- (a) **ball pen in combination with drinking vessel**: turn the vessel clockwise (switch focus to C) and push the pen button once
- (b) **three-button mouse**: push the right button
- (c) **keyboard**: push the right arrow key twice (switch focus to C) and hit the enter button once
- (d) **magnets** (three magnets put in a horizontal line): touch the rightmost magnet with the forefinger

3. *Push button B.*

- (a) **ball pen in combination with a cylindric object**: turn the cylindric object counter-clockwise (switch focus to B) and push the pen button once
- (b) **three-button mouse**: push the middle button
- (c) **keyboard**: push left arrow key once to switch the focus and push the enter button once
- (d) **magnets** (three magnets put in a horizontal line): touch the middle magnet with the forefinger

*

Basically, we can identify three patterns: (A) The user changed the focus by turning a cylindric object and confirmed the input by pushing a button, (B) the tester arranged three similar objects in a horizontal row (like the virtual counterpart, which is conform to PI) and consequently superimposed focus switching with task execution by simply touching an object, and (C) the participant used the arrow keys on a PC keyboard to switch focus; input is confirmed by hitting the enter button. These three patterns have been consistently used among the users.

Three patterns were
be observed.

B2: Radio Buttons

Three radio buttons (A, B, C) are arranged below each other. Selecting a button causes all others being deactivated.

Precondition: Focus is on radio button A.

Tasks and Patterns1. *Select item B.*

- (a) **magnets** (three magnets arranged in a vertical row): flip, grab, or touch the magnet in the middle
- (b) **ball pen:** point at radio button B on the screen and push the pen button

2. *Select item C.*

- (a) **magnets** (three magnets arranged in a vertical row): flip, grab, or touch the lowermost magnet
- (b) **ball pen:** point at radio button C on the screen and push the pen button

3. *Select item A.*

- (a) **magnets** (three magnets arranged in a vertical row): flip, grab, or touch the uppermost magnet
- (b) **ball pen:** point at radio button A on the screen and push the pen button

*

PI superimposed selection of a radio button with focus switching.

The patterns identified in this scenario have also been consistently used over the tasks. Like in the previous scenario, the magnets were arranged identical to the virtual widgets in the GUI. Thus applied PI superimposed focus change with radio button selection.

B3: Check Boxes

Two check boxes (A, B) are arranged below each other. Each box can be either selected or deselected.

Precondition: Focus is on check box A.

Tasks and Patterns

1. *Select check box A.*
 - (a) **stapler/puncher**: push it
 - (b) **mobile phone**: push a button on the numeric keypad
 - (c) **ball pen**: point at check box A on the screen and push the pen button
2. *Select check box B.*
 - (a) **stapler/puncher**: move it down and push it
 - (b) **mobile phone**: push the lower button on directional pad and push the same button as used for selecting A
 - (c) **ball pen**: point at check box B on the screen and push the pen button
3. *Deselect check box A.*
 - (a) **stapler/puncher**: move it up and push it
 - (b) **mobile phone**: push the upper button on directional pad and push the same button as used for selecting A/B
 - (c) **ball pen**: point at check box A on the screen and push the pen button

*

Again, the patterns observed have been consistently used over all tasks. Physical Instantiation has not been observed.

For a small number of UI widgets, PI

causes too much overhead.

We think that people consider rebuilding a GUI consisting of just two binary UI widgets unnecessary overhead.

Gestures for focus switching followed the layout of the UI widgets.

Focus switching was either done by object movement (up/down), by pushing buttons (up/down), or via pointing. Changing the focus thus seemed to follow the spatial layout of the UI widgets in the GUI: If they were arranged horizontally, the user also used horizontal movement to switch focus; vertical arrangement implied vertical focus switching.

Stapler and puncher both provided good haptic feedback.

Regarding stapler and puncher, the testers who used these objects stated that they provided a good haptic feedback.

*

For scenarios B1–B3, similar patterns were observed.

Comparing scenarios B1–B3, we can see similar gesture patterns. These tasks were primarily push-dominant as the participants tended to use button-equipped objects such as a keyboard, a mobile phone, a three-button mouse, and a ball pen, or they used objects looking like buttons, such as magnets. Of course, mouse and keyboard were not allowed to be used as normal but for appropriation, i.e., as primitive multi-button devices.

Similar methods were used for switching the focus, e.g., PI.

Focus change was either achieved with buttons, via PI (cf. above), or by dedicating a cylindric object as turn-to-switch-focus tool. Furthermore, some testers used objects representing a pointer, such as a pen or glue stick for pointing at the UI widget on the screen to activate its focus. Physical instantiation was applied by one third throughout these scenarios.

Most participants could handle the tasks using one object.

Consistently, for all three scenarios, eleven participants could handle the tasks with a single object. At maximum, three objects were needed for button and radio button but only two items for the check box scenario. Consequently, the number of objects used was at maximum as high as the number of UI widgets available in the GUIs.

B4: Rotary Buttons

Two rotary buttons (A, B) with selection indicators are arranged next to each other. Each widget can pick a value between 0° and 359° .

Precondition: Focus is on rotary button A. Both indicators point at twelve 'o clock position.

Tasks and Patterns

1. *Set B to three 'o clock.*
 - (a) **two drinking vessels/bottles/caps** (arranged next to each other): turn the right object approximately 90° clockwise.
 - (b) **two ball pens** (arranged in parallel next to each other): turn the right pen approximately 90° clockwise around its center of gravity
2. *Set A to six 'o clock.*
 - (a) **two drinking vessels/bottles/caps** (arranged next to each other): turn the right object approximately 180° clockwise.
 - (b) **two ball pen** (arranged in parallel next to each other): turn the right pen approximately 180° clockwise around its center of gravity

*

Not surprisingly, this scenario was turn-dominant. In total, 15 participants used round-shaped objects, such as drinking vessels, bottles or magnets that suggested turning for controlling the soft rotary buttons.

As regards quantity of object usage, 15 testers used two items to solve the tasks. Ten out of these used nearly identical objects for rebuilding the UI, therefore eliminated the need for explicitly changing the focus between the

This scenario was turn-dominant.

Ten users rebuilt the GUI with two similar-looking objects.

Again, continuous input is controlled via PI.

widgets by directly interacting with the physical artifacts. Where available, physical idiosyncrasies, such as the handle of a cup, were exploited as value indicators.

Besides, four testers used pens for solving the tasks using them as clock hands by putting the tip to the desired position, i.e., three 'o clock or six 'o clock, respectively.

Unlike the check box scenario in which PI for just two UI widgets was considered too much overhead, the paradigm was applied for the two rotary buttons. For widgets of continuous input, PI was perceived to be easily applied when the dedicated input device is missing (cf. 4.4.5—“Qualitative User Feedback”). The patterns represented turn-intensive gestures manipulated on cylindric-shaped objects that suggested turning. The patterns have been consistently applied over all tasks.

B5: Sliders

Three horizontal sliders (“Red”, “Green”, “Blue”) are arranged below each other. Each Slider adjusts the value (0–255) of a color channel (red, green, or blue).

Precondition: Slider focus is on slider “Red”.

Tasks and Patterns

1. *Pick the middle of the slider range from slider “Red”.*
 - (a) **three magnets** (arranged below each other, put on the left side of the table): move the uppermost magnet to middle of the table.
 - (b) **three whiteboard markers** (standing or lying flat below each other, put at the left side of the table): move/roll the uppermost marker (= red color)

2. *Pick a value between the middle and the end of the slider range from slider "Blue".*
 - (a) **three magnets** (arranged below each other, put on the left side of the table): move the lowermost magnet a bit farther than to the middle of the table.
 - (b) **three whiteboard markers** (standing or lying flat below each other, put at the left side of the table): move/roll the lowermost marker (= blue color) farther than to the middle of the table
3. *Pick value "255" from slider "Green".*
 - (a) **three magnets** (arranged below each other, put on the left side of the table): move the magnet in the middle to the right table end.
 - (b) **three whiteboard markers** (standing or lying flat below each other, put at the left side of the table): move/roll the marker in the middle (= green color) to the right table end

*

In the slider scenario, PI was identified strongest among all scenarios. Eleven participants used this paradigm (cf. Figure 4.7). In most cases, magnets were arranged in a vertical row and moved from left to right to adjust the corresponding color channel in the GUI. Alternatively, the users utilized marker pens and rolled or moved them from left to right.

As regards the sliders, PI was applied most often among all scenarios.

Interestingly, for the second pattern, the testers mapped the colors of the whiteboard marker pens (red, green, and blue) to the sliders for the color channels in the GUI. Color mapping was also observed in the jump and run scenario, in which some testers mapped the action of firing to a red button. Besides the patterns described above, two participants used the whiteboard and corresponding color markers to draw the slider trace. Interestingly, they arranged the slider traces from top to bottom according

Users mapped the color of a marker pen to the corresponding color of the slider.

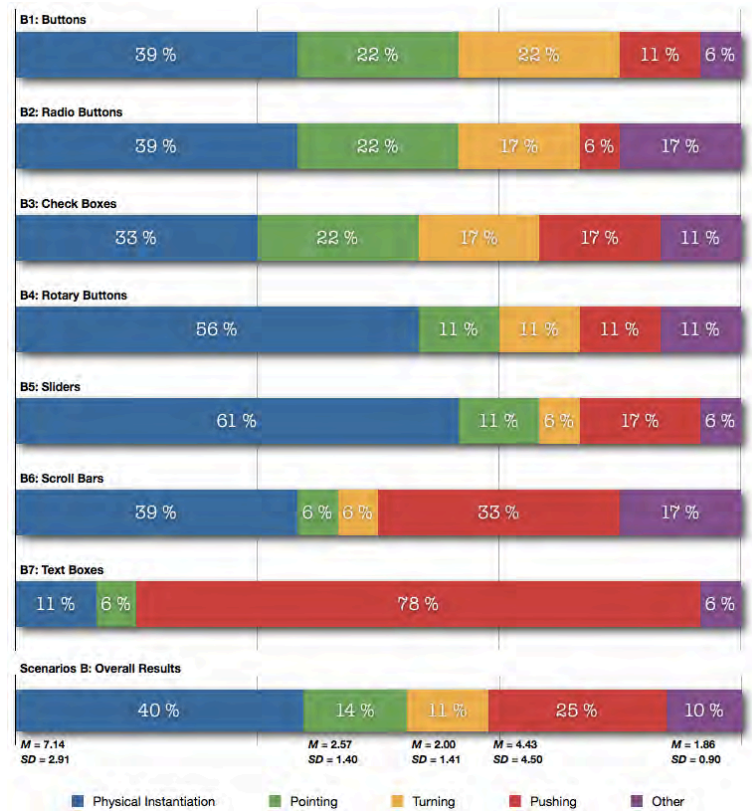


Figure 4.7: Methods for switching focus between multiple instances of a UI widget per scenario applied by the users ($N = 18$). Application of PI was most frequently observed for continuous input controls (B4, B5). Scenarios B1–B3 showed similar patterns for focus change. As most users utilized the keyboard for typing characters in the text boxes (B7), they also used key-based navigation for focus switching. However, using the keyboard as usual was not conform to the idea of appropriation. Moving an object for changing the focus, for instance, is denoted by “other”.

to the *sequence of tasks*, i.e., they drew a red line first, then a blue line in the middle, and finally, a green line at the bottom. In the GUI, however, the sliders were arranged as follows: red, green, blue (top to bottom).

Most participants solved the task using three objects (eleven people) or less (three people used one object each, four people utilized two items each).

B6: Scroll Bars

Two columns (A, B), ten rows (1–10), and a horizontal and a vertical scroll bar are composed to a spreadsheet interface.

Precondition: The table is scrolled to the top. Cell (A,1) has the focus.

Tasks and Patterns

1. *Scroll to the very bottom of the table.*
 - (a) **rectangular underlay/sheet of paper combined with ball pen/finger:** move the pen/finger from the upper right corner of the underlay/paper to the lower right corner.
 - (b) **keyboard:** push the arrow down key a few times
2. *Select cell (B,5).*
 - (a) **rectangular underlay/sheet of paper combined with ball pen/finger:** tap with the pen/finger on the underlay/paper relatively to the cell position on the screen.
 - (b) **keyboard:** push the arrow down key four times and push the arrow right key once
3. *Select cell (A,2).*
 - (a) **rectangular underlay/sheet of paper combined with ball pen/finger:** move the pen/finger from the lower right corner of the underlay/paper to the upper right corner and tap with the pen/finger on the underlay/paper relatively to the cell position on the screen.
 - (b) **keyboard:** push the arrow up key three times and push the arrow left key once
 - (c) **mobile phone:** push the upper button and the left button once on the directional pad

4. *Scroll the table to the very right.*

- (a) **rectangular underlay/sheet of paper combined ball pen/finger:** swipe the pen/finger to the right on the bottom edge of the underlay/paper
- (b) **keyboard:** push the arrow right key a few times
- (c) **mobile phone:** push the right button on the directional pad a few times

*

One third of the users used key-based cell navigation.

The scroll bar/spreadsheet scenario was push-, move/swipe- and tap/touch-dominant. One third of all participants used key-based navigation on a mobile phone or a PC keyboard for incrementally switching the focus between the spreadsheet cells. Implicit scrolling of the table was expected for cells selected outside the visual scope of the table.

As regards PI, a flat and rectangular object represented the virtual table.

Opposed to this appropriation principle, seven testers applied PI (cf. Figure 4.7), using a flat and rectangular auxiliary item, such as a sheet of paper that represented a physical instance of the table. By moving or swiping with the fingers or an object, such as a pen or a glue stick, on the physical instance, users managed to scroll the table, whereas tapping/touching directly selected a single cell.

Interestingly, most users just needed one object for solving the tasks.

Despite complexity, most testers solved the tasks using one object (13 people in total; two objects were used at maximum). This was due to the fact that the testers either operated a multi-button device or used their fingers for tapping on another object.

The spreadsheet interface can be considered maximum acceptable complexity for IUIs.

Being considered neither easy nor difficult for finding objects and gestures for appropriation in this scenario (Q6, Q7), we can conclude that this is an example for the maximum complexity of an IUI accepted by participants. In fact, this scenario was perceived least easy among all A- and B-scenarios. What we can conclude from this scenario is that people were still able to find solutions for solving the tasks, but appropriation reaches its limits for interfaces of the types of complexity.

B7: Text Boxes

Two text boxes (A, B) are arranged below each other.

Precondition: Text box A has the focus.

Tasks and Patterns

1. *Write/type "Hello" in box B.*
 - (a) **keyboard:** hit the tab key or the arrow down key and hold the shift key while pushing the "h" key, followed by pushing "e", twice "l", and once "o"
 - (b) **whiteboard marker pen:** draw two squared boxes below each other and write "Hello" in the second box
2. *Write/type "2011" in box A.*
 - (a) **keyboard:** hit the tab key or the arrow up key, push "2", "0", and twice "1"
 - (b) **whiteboard marker pen:** write "2011" in the upper box

*

For the text box scenario, we observed the most consistent interaction pattern regarding the usage of objects and gestures. Not surprisingly, 14 people solely used the PC keyboard for both, changing the focus between the text boxes and for typing the demanded characters. However, some people found other options, e.g., they used T9-like input via the numeric keyboard of the mobile phone, or they used white board marker pens to draw the text boxes on the white board and wrote the demanded characters directly into these rectangles.

Not surprisingly, most users grabbed the keyboard for solving the tasks.

The text box scenario was the one that did not contribute to intended appropriation of everyday objects, as rather dedicated input devices (keyboard, pen) were

As regards writing or typing, appropriation was no feasible option.

used. Yet, we can derive two important findings from this scenario: (1) As regards writing or typing, people basically find no feasible solution in terms of appropriation; text input via non-dedicated devices is inappropriate. (2) As most participants have used dedicated devices for text input, finding objects and gestures was – of course – perceived easiest in this scenario. These values (finding objects: $M = 4.67$, $SD = 0.59$, finding gestures: $M = 4.56$, $SD = 0.70$) can be considered maximum satisfaction in the scope of this study.

4.4.5 Qualitative User Feedback

Feedback was collected with the aid of a questionnaire.

Besides objective feedback collected from the video recordings, we asked the users for their subjective impression on IUIs as experienced in this study. The feedback has been collected by questionnaires handed in after the last scenario had been completed. The full feedback form is printed in A—“Appendix for the Wizard of Oz Study”; references to the corresponding questions including mean and standard derivation are put in brackets. Unless otherwise noted, the given scores are averaged values from a five point Likert scale, 1.0 meaning “strongly disagree”, 5.0 meaning “strongly agree”.

The users considered appropriation of objects for controlling digital interfaces a convincing concept despite disbelief in technical realization.

Principally, the testers agreed that appropriation of passive objects for controlling digital interfaces is a convincing concept (Q1: $M = 4.11$, $SD = 0.83$). However, some participants doubted technical feasibility regarding implementation of the WOZ system, such as the anticipated interpretation of the user’s gestures applied on the objects. Disbelief in technical realization was confirmed by the users as they thought that – for a functioning IUI – it would be difficult for the system to distinguish between voluntary and non-voluntary input with (appropriated) objects (Q2: $M = 4.00$, $SD = 0.77$). Nevertheless, the participants suggested solutions to this problem: Most people said that pointing at the system or any special activation gesture could be a solution for expressing intended appropriation. Besides, some testers proposed to allow appropriation only to be recognized within in a specific area, e.g., in front of the remote target system.

Regarding the choice of objects, textiles, such as a pillow, a piece of cloth, clothes, shoes and a backpack were considered inadequate for appropriation. These objects are too soft. The cable was also regarded inadequate for this reason (Q3).

Textiles were regarded too soft for appropriation.

In some cases, appropriation of objects led to unexpected behavior (Q5). This was observed for objects that contained liquid or gas. In cases where the substance could easily escape, e.g., by pushing the button of a deodorant spray, appropriation led to surprise as the user did not consider possible substance escape. For objects that needed intense physical effort to open them, such as a water bottle, where the cap needs to be turned multiple times until it is open, no surprise was observed.

As regards some objects, chemical substances could escape while being appropriated.

Concerning object properties that the testers considered suitable and necessary for appropriation, people preferred lightweight and small objects that should be equipped with physical controls rather than have a smooth surface (Q4, cf. Table 4.2). This could be confirmed by the video observations, as people preferred having plastic items over glass (bottle, drinking vessel, etc.) and a small mobile phone over a stationary phone. As regards object deformation, no clear preference could be identified, i.e., 50% preferred rigid objects over flexible objects (and vice versa) and 50% expressed no preference at all. However, as mentioned above, for those who preferred flexible surfaces, the material must not be too soft.

Users preferred small and lightweight objects for appropriation.

Regarding physical properties of objects, these seem to be more important for A-scenarios (Q11: $M = 4.11$, $SD = 0.83$) compared to B-scenarios (Q16: $M = 3.61$, $SD = 1.04$). This may be due to the fact that objects appropriated in A-scenarios replaced physical controls, such as buttons, whereas in B-scenarios only virtual controls which have no haptic feedback, had to be controlled.

Physical object properties played a more important role for A-scenarios.

When being asked whether they preferred appropriation of one object using it as multi-purpose device or whether they preferred to appropriate multiple single objects (i.e., using one object to control one function), figures for A- and B-scenarios were roughly the same: 50% preferred

Q4: I preferred choosing objects which were/had...			
(a)	flexible/deformable 4	rigid 6	no preference 8
(b)	buttons/controls 14	smooth 1	no preference 3
(c)	lightweight 14	heavy –	no preference 4
(d)	small 11	large –	no preference 7

Table 4.2: Object properties preferred for appropriation. Users preferred small and lightweight objects equipped with buttons or controls. As regards rigidity/flexibility, no preference could be identified.

Independent of A- and B-scenarios, most users preferred appropriating one object as multi-purpose device for solving a scenario.

having just one object at hand that can be appropriated to execute many actions (Q10, Q15). This was consistent with our video observations as most people could handle all tasks per scenario with just one or two objects. Preference for multiple distinct objects was half as much compared to multi-purpose functionality. Thus, basically, we identified two groups of users, one preferring multi-purpose appropriation, and – a smaller number – one group that preferred exact matching of one object to one function. Figure 4.8 shows the complete results.

In general, finding suitable gestures for appropriating an object felt easier than picking a suitable object first.

While the testers were looking for objects suitable to be appropriated for a A-scenarios/tasks, they stated that they searched for objects that would actually be within reach if they encountered this scenario as a real setting (Q9, cf. Figure 4.9). This showed that the testers tried to identify themselves with the given scenario. In this context, we could see that finding suitable objects for appropriation felt easier for B-scenarios compared to A-scenarios (Q6, Q12, cf. Figure 4.10). Taking the ease of finding appropriate gestures into consideration, the participants generally indicated, that once an object had been found, applying a gesture felt easier than finding the right object itself (cf. Figure 4.10). We attributed this result to the physical affordances of the object, i.e., that they led the testers in finding appropriate gestures.

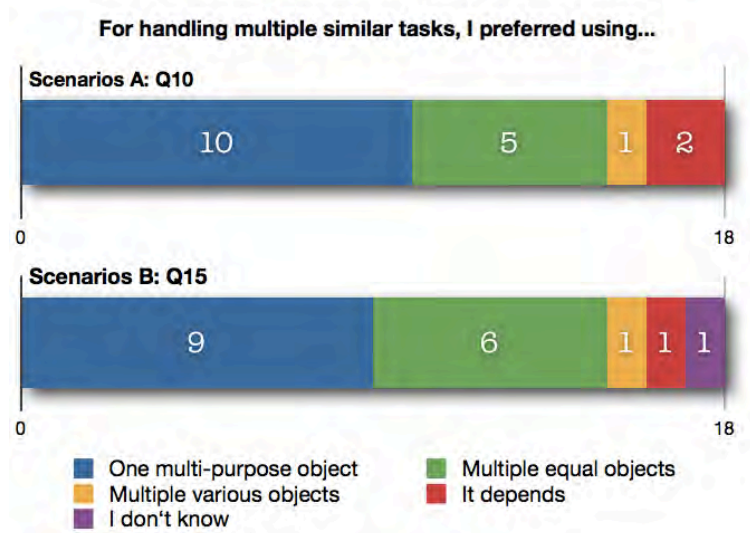


Figure 4.8: Preferred quantity of objects for solving a scenario. Most users ($N = 18$) preferred appropriating one object as multi-purpose device for solving all tasks of a scenario. A smaller group preferred utilizing multiple but equal/similar objects. In general, preference for quantity of objects was independent of A- and B-scenarios.

Regarding A-scenarios, operating the light and interacting with a virtual shop window have been perceived least easy in finding objects and gestures (all in the range of 3.22 and 3.94, thus below “agree”-level on the Likert scale, cf. Figure 4.10). From our video observations we could identify that for the first case, testers found it challenging to let the system know which of the two light sources they wanted to manipulate (focus switching). For the latter case, flipping the sides of the glasses demanded the participants to inspect the object(s) thoroughly again for finding either an adequate gesture or a new object plus gestures for solving the task.

As regards scenarios A, operating the light and interacting with a virtual shop window were perceived least easy.

Concerning B-scenarios, buttons, radio buttons, and check boxes nearly share the same results. Indeed, similar interaction patterns could be observed for these scenarios. Slider and rotary button scenarios were perceived easiest for finding objects and gestures. For these scenarios, PI was applied most often. As this paradigm is a specialization of

**I have chosen objects which would also be available in reality
Scenarios A: Q9**

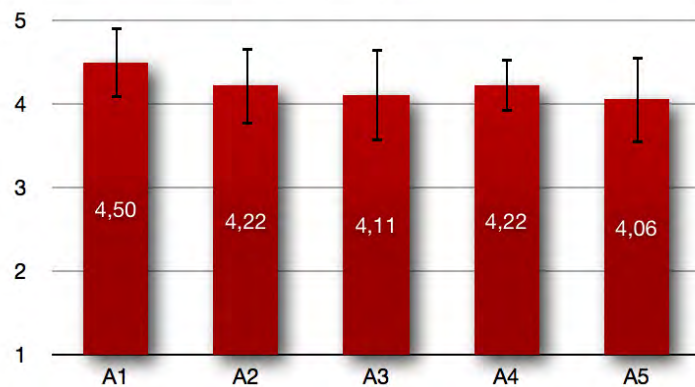


Figure 4.9: Match between selectable objects and objects available in reality. Overall, the testers agreed, that they had chosen objects that would be available if they encountered the scenario (A1–A5) in reality. A score of 1.0 means “strongly disagree”, whereas 5.0 means the opposite. Error bars indicate the 95% confidence interval for the mean.

For scenarios B4 and B5 in which PI was applied most often, finding gestures for solving the tasks was perceived easiest.

Norman’s natural mappings, we interpreted PI as a natural way and good choice for appropriating objects in order to use them as controllers for continuous input. The scroll bars scenario with the spreadsheet interface was perceived least easy. Although PI was applied here by several people, they stated that the interface felt complex compared to the other GUI mockups. In fact, this GUI consisted of two scroll bars and a spreadsheet that had to be scrolled in order to access all cells. Scoring 3.00 ($SD = 1.24$) on average concerning the ease of finding objects and the ease of finding gestures being rated 3.50 ($SD = 1.25$) on average, we can say that this GUI illustrates maximum complexity that can be handled through appropriation. Therefore, IUIs cannot help spontaneously controlling GUIs consisting of multiple different UI widgets. On average, finding objects for controlling the text box scenario was perceived easiest. This was no surprise as most testers took a keyboard for text input. This gives us a maximum value for the ease of finding objects and shows another border that IUIs cannot cross: Text input via appropriated objects is no satisfying solution.

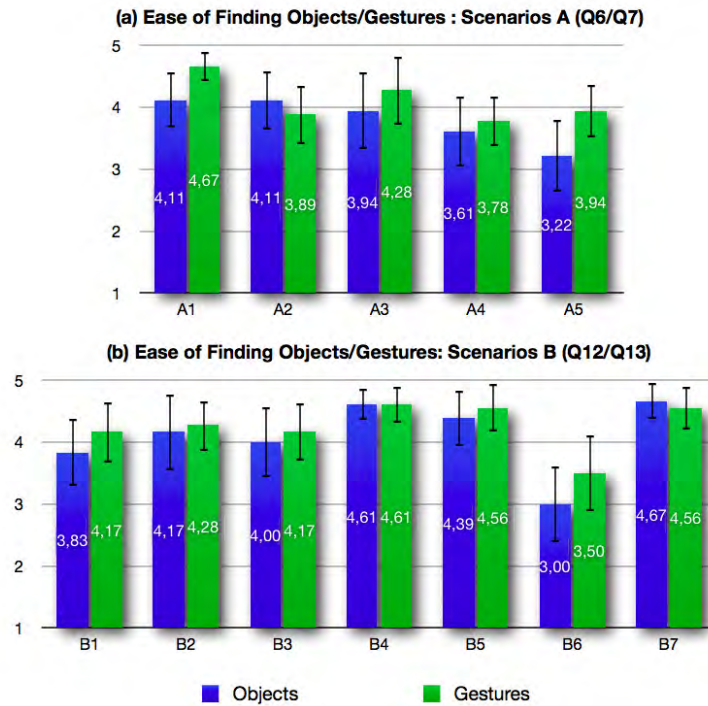


Figure 4.10: Perceived ease of finding objects and applying suitable gestures on them. The bars represent the average rating on a five-point Likert scale, 1.0 meaning, that the users strongly disagreed that finding objects/gestures was easy for the corresponding scenario, whereas 5.0 means the opposite. Error bars indicate the 95% confidence interval for the mean.

4.5 Lessons Learned

The WOZ study helped us to gain insight into the mental models of our participants regarding appropriation of everyday objects for IUIs. Generally, we took note of the following findings:

- For scenarios A, i.e., remote target systems, people principally tried to *imitate the missing input device* through button-equipped everyday objects.

- Shape was probably the most important object property for deciding which object to pick. People tended to imitate either the shape of the input controller or the target that was to be controlled.
- Besides, some people looked for colors on objects that could be associated with a certain action.
- Pointing was a natural way of activating a system. It was used to indicate that object manipulation should have an effect on the system that was being pointed at.
- Generally, one third of the testers tried to rebuild a GUI with objects by looking for similar shapes and by adopting the layout of the UI widgets. Focus switching and manipulation were superimposed.
- Physical instantiation was more likely to be applied in case continuous input was expected, such as spatial movement, rotation, or slider navigation.
- In most cases, people can handle a scenario with just one object. People who needed more objects to solve a scenario used one object per task at maximum; multiple simultaneous object use was rare. These observations were not valid in cases where PI is applied, as rebuilding the UI generally required as many objects as virtual UI widgets were presented.
- For interaction with GUIs, in case PI was not applied, people used either a special object or a special gesture to sequentially switch the focus from one UI widget to another.

A complete comparison for focus switching solved by the testers is given in Figure 4.7.

A higher ranking for an object in the diary study did not correlate with frequency of appropriation.

Comparing the results of the diary study with object usage during the second study, we can't see a correlation between frequency of objects within reach (diary study) and the frequency of usage of objects as observed in this study. For example, the pen was used most often. In the diary study, however, it was ranked sixth. On the other hand, the bottles were used less frequently than the pen in the WOZ

study but ranked higher in the diary study. So basically, if an object is generally “more omnipresent” among users, this does not imply that it will also be used frequently for appropriation. Figures 4.11 and 4.12 illustrate all objects used by the participants including the applied gestures as a summary.

Having identified hypothetical interaction patterns for IUIs, the next step is to investigate interaction based on these patterns with an interactive prototype featuring real time feedback. Furthermore, the interplay of normal object usage and appropriation will be evaluated in the final user study (cf. 5—“Testing Everyday Object Interaction with an Interactive Prototype”).

Next, appropriation will be tested with an interactive prototype.

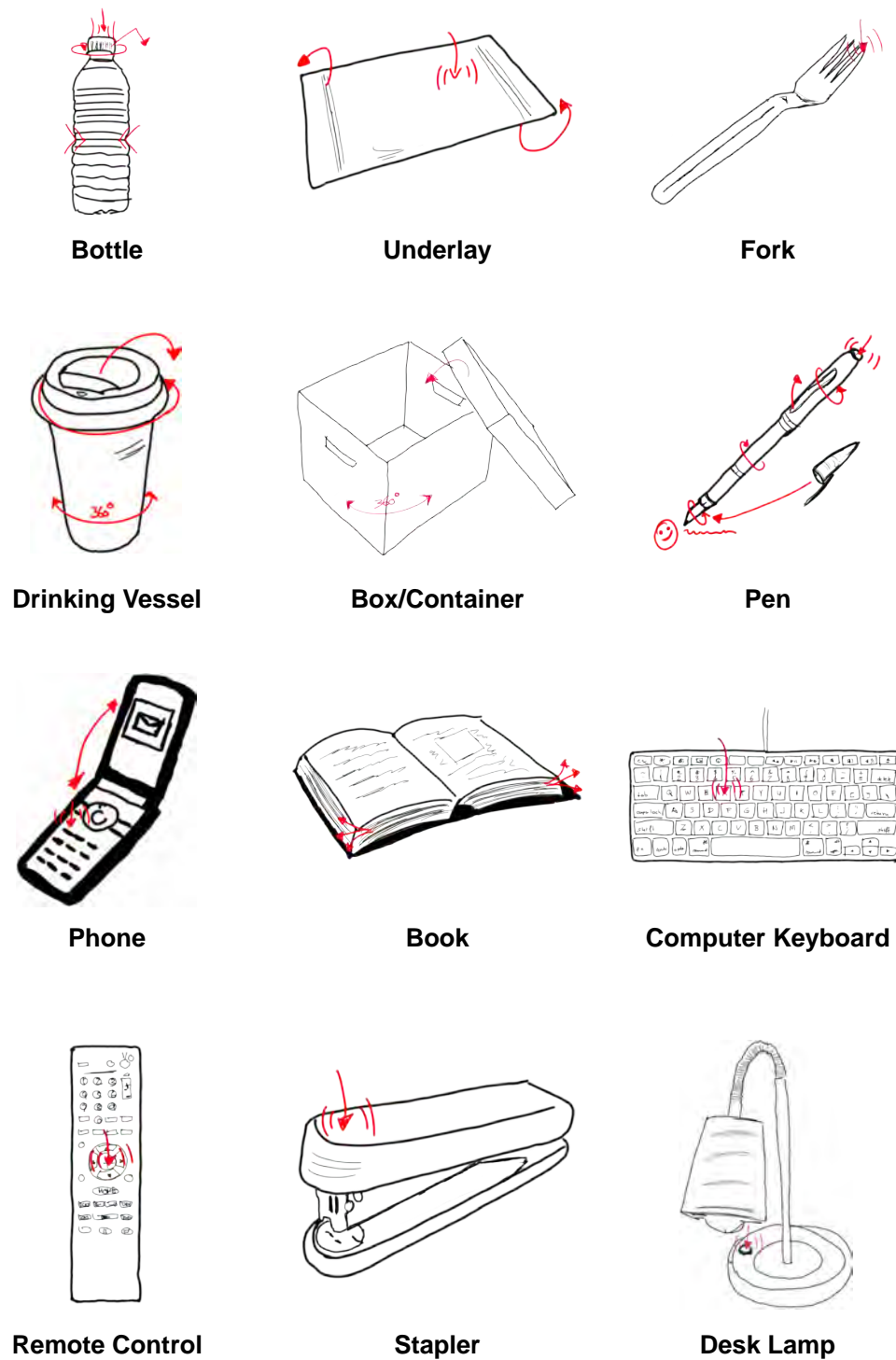


Figure 4.11: Everyday objects identified by our diary study and used in appropriated in the WOZ study (1/2). Red arrows show actual manipulation behavior as identified in the WOZ study; the object was at least used once in this way by the testers.

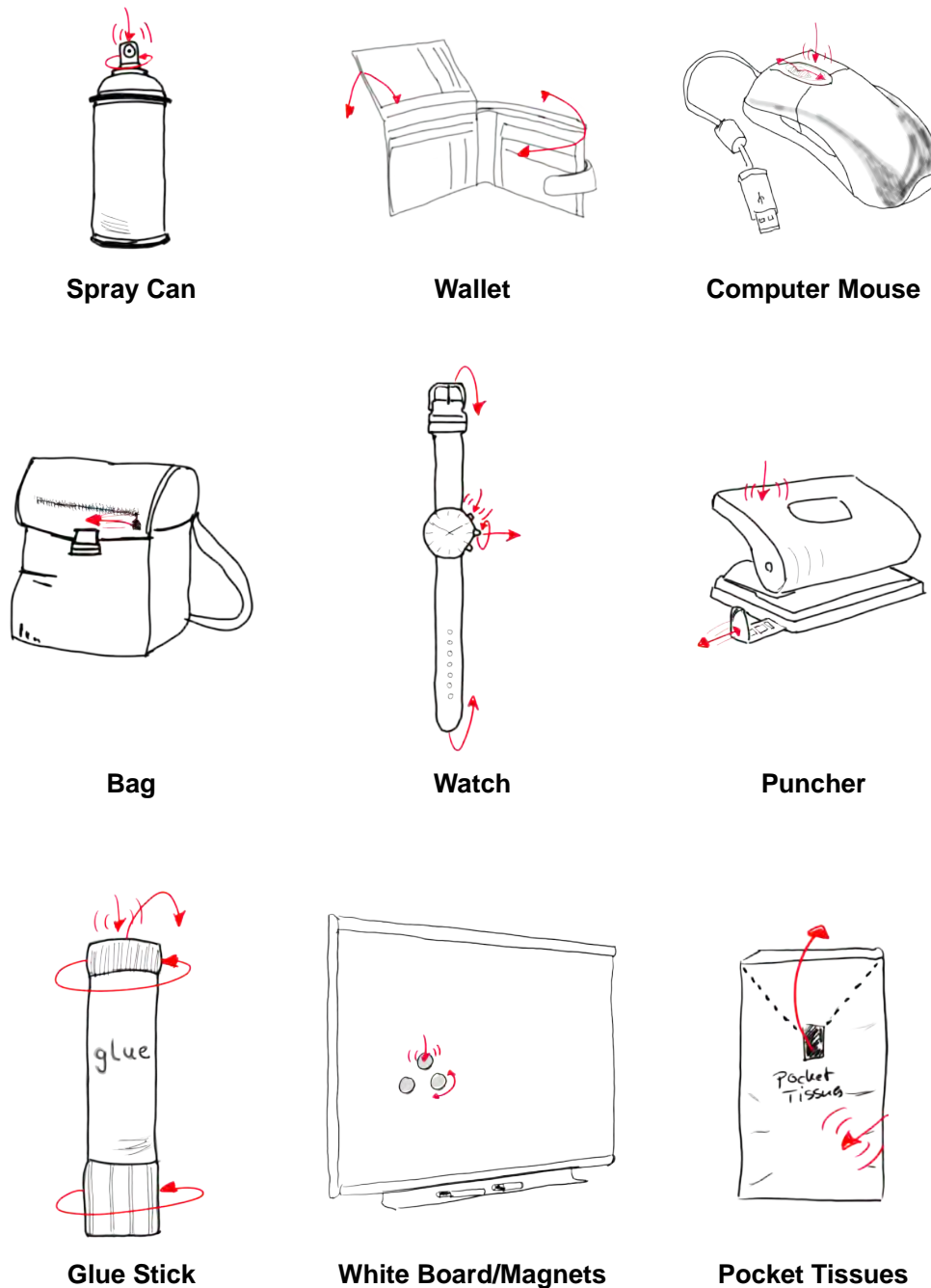


Figure 4.12: Everyday objects identified by our diary study and used in appropriated in the WOZ study (2/2). Red arrows show actual manipulation behavior as identified in the WOZ study; the object was at least used once in this way by the testers.

Chapter 5

Testing Everyday Object Interaction with an Interactive Prototype

Reflecting the previous user study, our primary goal was to understand how users manipulate everyday objects in terms of appropriating them as controllers for daily encountered user interfaces, such as televisions or ceiling lights (cf. 4—“Wizard of Oz Study: Interaction Patterns for Instant User Interfaces”). Although we gained insight into people’s mental models for a limited set of objects in selective scenarios, the WOZ prototype lacked real time feedback: The system state was not updated until a task was fully completed. According to Norman [2002], direct feedback dramatically influences the user’s mental model, i.e., her interpretation regarding how the system works.

Direct feedback influences a user’s mental model dramatically.

In addition, the previous study merely concentrated on finding suitable gestures for appropriation by exploiting physical affordances. However, this does not reflect a realistic scenario in which the user would mix normal object usage with appropriated use. Testing everyday object interaction under realistic circumstances requires to move away from WOZ-style experimentation and move on to an interactive prototype.

Testing the interplay of normal object usage and appropriated use requires a prototype with real time feedback.

This chapter presents an interactive prototype that was tested by users.

In this chapter we present the design of an interactive prototype which is tested by users in two scenarios covered by a qualitative study. Based on the goals of this study and the results of the preceding WOZ experiment, we present design implications for the prototype including a technical description. We present the design of the study and its procedure followed by the evaluation of the results and a conclusion.

5.1 Goals

The interactive prototype was built for giving users a realistic impression of IUIs.

The main purpose of the interactive prototype is to give the user a realistic impression about IUIs. Besides an authentic prototype, the design of a realistic scenario in which the prototype is to be tested is also important. For everyday object interaction in the sense of IUIs, normal object usage (e.g., drinking from a bottle) will be mixed with the new interaction metaphor (e.g., turning the bottle for controlling a dimmable lamp).

An IUI must know when to interpret object manipulation and when not.

The IUI system must understand when interpretation of object manipulation is desired and when it is not: Imagine a user who screwed a bottle cap for controlling the volume of a TV (*mode of appropriation*). Later, she might want to drink from the bottle which requires turning the cap again. However, in this case she obviously does not want to influence the volume level; hence, the system should not interfere.

A user can uncouple an object either explicitly or implicitly form an IUI for letting the system know that object manipulation should not be interpreted.

In this user study, we concentrated on how this interplay affected on the user experience by testing two different methods for unlinking an object from the IUI system. The first method is based on a pointing gesture for *explicitly* letting the user deactivate an object from appropriation mode. The second variant imitates a hiding metaphor: The user can either unregister the object *explicitly* through hiding by holding the palm over it for a few seconds, or the artifact is *implicitly* deactivated when the user puts the object to the side. Regarding the first method, the user is completely responsible for deactivating the object, which gives her full control about the state an object is in.

The second variant basically enables the user to be still of control of the object state. Yet, the system can interfere and reduce the burden from the user.

As regards these deactivation principles, we wanted to find out:

- Which method is preferred?
- What patterns can be observed regarding object activation and deactivation by the users? Do they, for example, tend to immediately deactivate an object after it is used, or are they not concerned about deactivating an artifact before normal object usage is desired?
- Which interaction problems do occur regarding both variants?

Testing the interplay is different from the iCon user study by Cheng et al. [2010], in which the authors used a purely time-based activation and deactivation metaphor (cf. 2.2.8—“iCon: Cheng et al., 2010”). A mixture of normal usage and appropriated use was not tested.

Besides, we wanted to test further qualitative aspects towards everyday object interaction. As we have learned how people appropriate artifacts in given scenarios (WOZ study), we tried predefined object-action-gesture mappings in this study. This approach removes the necessity of end user programming and allowed for *instant* interaction. Yet, the question is whether such predefined mappings are a feasible solution, or whether they lead to confusion. Finally, we wanted to know whether everyday objects are considered a valuable and comfortable solution assuming that the dedicated input device is missing. Could users image utilizing everyday objects in such situations?

Our aim was to investigate qualitative aspects regarding user experience with IUIs.

Before finding answers to these questions, we describe the prototype design. Study design, conduction, and evaluation are followed up.

5.2 Prototype Design

Strictly speaking, a system providing access to IUIs has two major responsibilities:

1. It must turn any *passive* (everyday) object into an *interactive* controller through gesture recognition.
2. It must transform recognized gestures into events which are forwarded to the target system, i.e., the interface that is to be controlled.

5.2.1 Breathing Life into Passive Objects

Instant access to IUIs implied minimal object preparation.

The choice of technology for turning passive objects into active controllers depends on the design decisions. As regards Co-Optjects as IUIs, we were interested in giving the user *instant* access to appropriation of artifacts. This requires minimization of any object preparation before it can respond to gestures. Hence, choosing a technology that requires the user to attach variant physical sensors to the object is unacceptable.

As regards gesture recognition, we used a depth camera combined with marker detection.

Besides instant access, the type of gestures that have to be recognized influence the choice of technology. In the WOZ study we mainly identified (1) touching/pushing, (2) planar movement, (3) stationary turning, and (4) pointing. As regards (1), we decided to take a depth camera approach that makes any surface – regardless of being rigid or flexible – sensitive for touch. Gestures of category (2), (3), and (4) were detected through marker recognition. These markers were attached to the physical artifacts for being captured by a web camera that delivered precise position and angle of an item.

For touch detection, no object preparation was needed.

Attaching paper fiducials is a quick approach and therefore only slightly delays instant access. For pure touch- and pressure detection, the markers were not needed.

Touch Detection on Arbitrary Surfaces

Investigating touch detection for specifically prepared surfaces has been widely anticipated in research. Examples are multi-touch tables based on FTIR (Frustrated Total Internal Reflection) [Han, 2005], DSI (Diffused Surface Illumination) [Akechi et al., 2011] or capacitive touch screens as often embedded in mobile devices such as the [Apple iPhone](http://www.apple.com/iphone)¹. However, these multi-touch devices are specifically designed and therefore cannot be used for breathing life into everyday objects without altering object-specific properties, such as shape, material, or weight.

Standard touch detection technology is impractical for making existing objects responsive to touch.

In addition, these technologies neglect deformable structure. Everyday objects come in various shapes with rigid or flexible surfaces. A feasible solution to this problem is the use of depth cameras. Having become affordable thanks to mass production, experimentation with the [Microsoft Kinect](http://www.xbox.com/kinect)² depth camera has risen interest by researchers and allows to make any surface touch-sensitive.

Depth cameras allow for ad hoc touch detection on arbitrary surfaces.

For our prototype, we created an aluminum mounting (height: 70 cm) as depicted in Figure 5.1. The Kinect was fixed to the top of the mounting facing the workspace beneath it. The Kinect depth camera (cf. Figure 5.2) consists of an RGB camera with a resolution of 640x480 pixels² and a frame rate of 30 fps. Moreover, the chassis embeds an infrared projector and an infrared camera. In order to determine the depth of an image, the projector emits light pulses that are reflected and captured by the IR camera. The distance between each pixel in the scene and the camera is retrieved by the time each pulse needs from emission until detection [Borenstein, 2011]. Using infrared light for depth scanning is fast compared to traditional laser scanning, as the Kinect emits the light with a single shot. Laser techniques are time-consuming and therefore interfere with the fluidity of interaction since the scene needs to be sampled. Another advantage of IR light over traditional depth scanning is, that the light is invisible to human eyes and thus does not influence the user. Similarly to heat maps as known from thermographic

Depth scanning through the emission of IR light is fast compared to traditional laser scanning.

¹www.apple.com/iphone

²www.xbox.com/kinect

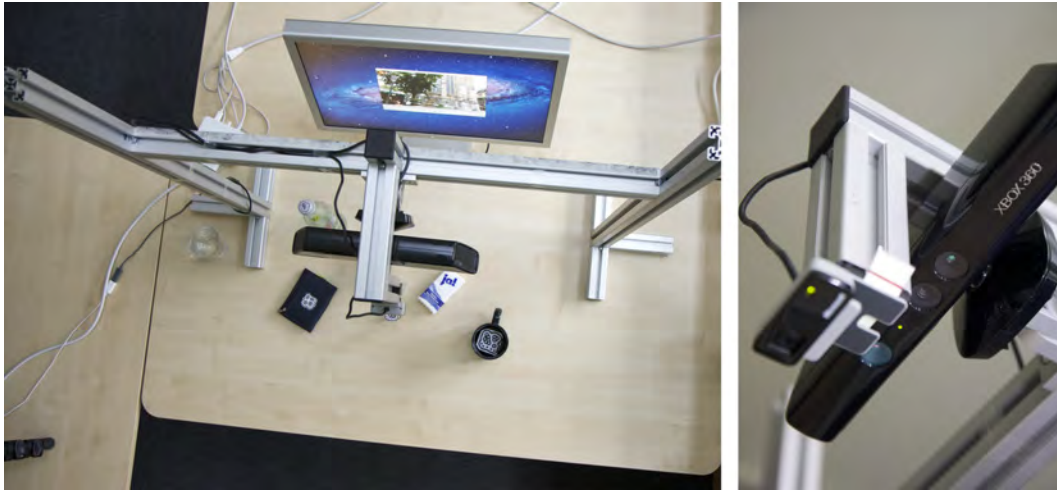


Figure 5.1: The prototype aluminum mounting. Left: The mounting photographed from bird's eye view. The Kinect depth camera (black bar in the middle) is facing the table surface and captures the top surfaces of the everyday objects put on the table. Behind the mounting, we put a monitor. Right: Close-up view showing the Kinect and the web camera. Both cameras share almost the same perspective.

cameras, depth maps can be visualized with colors. Figure 5.3 shows an example. Reddish parts are closest to the camera, whereas greenish parts are farthest away. The following example shows how depth information can be used for touch-/pressure detection on arbitrary surfaces.



Figure 5.2: Kinect teardown. Basically, the Kinect consists of three components (from left to right): an IR projector, an RGB camera, and an IR camera. (Image: iFixit.com)



Figure 5.3: Colored depth map. This picture has been taken by a depth camera (Kinect). Reddish parts appear closest to the camera, whereas greenish parts are farthest away.

TOUCH DETECTION THROUGH DEPTH CALCULATION (1/2):

Assuming the Kinect is positioned one meter above a flat table surface, the corresponding depth image delivers a depth value of 100DU (DU: Distance Unit) per pixel. Putting a pack of pocket tissues on the desk, the depth values for those pixels representing the tissues will change. The pack has a height of 10 HU (HU: Height Unit) and therefore will appear closer to the camera than the uncovered table surface. Assuming it holds $1\text{DU} \equiv 1\text{HU}$, the distance from the top surface of the tissues to the camera is $100\text{DU} - 10\text{DU} = 90\text{DU}$.

Whenever something is approaching the tissue surface, e.g., a finger, the depth values for those pixels occluded by the finger will change as did the depth values change likewise to the tissues put on the table. Figure 5.4 depicts the situation including the corresponding depth values. As regards the tissues, pressure, touch, or hover are detected whenever the distance to the camera differs from 90DU. A depth value larger than 90DU would mean that the tissues are being *pressed*, thus flattened and consequently farther away from the camera. A depth value smaller than 90DU would mean that something is put on the tissue surface, e.g., a finger.

Example:

*Touch Detection
through Depth
Calculation (1/2)*

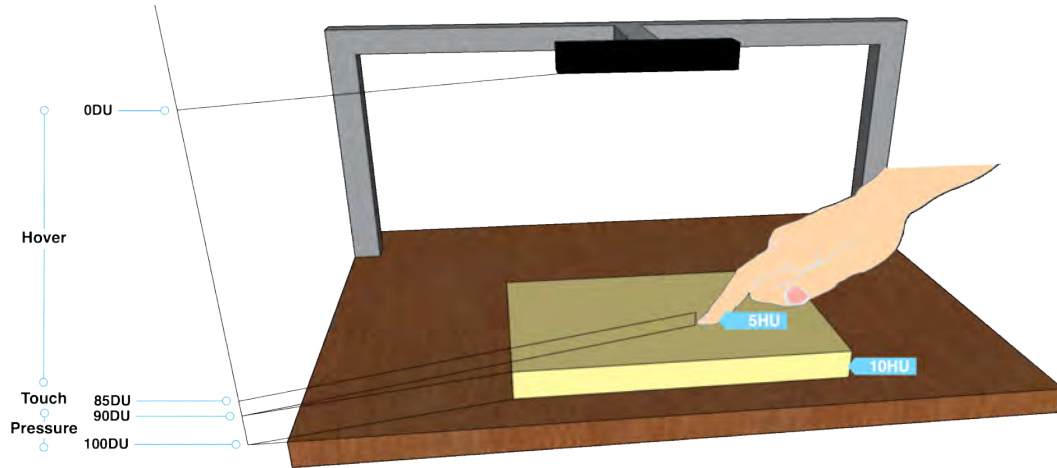


Figure 5.4: Touch detection through depth calculation. The distance from the camera (black bar) to the table surface (brown) is 100HU. As the pack of pocket tissues (yellow) has a thickness of 10HU, it appears closer to the camera than the table. When a finger touches the tissue surface, it is 85DU away from the camera. Pushing the tissues means that the finger is farther away from the camera, whereas hover is detected when the finger is closer to the camera.

TOUCH DETECTION THROUGH DEPTH CALCULATION (2/2):

Example:
Touch Detection
through Depth
Calculation (2/2)

In order to prevent hover effects, like when the user is putting the finger towards the tissues before actually touching them, or in order to sense different intensities of pressure, we can define depth ranges. Assuming the height of the table, the pocket tissues, and the finger tip to be as given in Figure 5.4, a touch is recognized when the depth value of the pocket tissues (around the finger tip) is around $d_{table} - h_{tissues} - h_{finger}$, or – given in numbers – $(100 - 5 - 2)DU = 93DU$ (d_x represents a depth value, whereas h_x represents a height value). As regards pressure, we yield a range of $(d_{table} - h_{tissues} - h_{finger}, d_{table} - h_{finger}]$, i.e., $(93, 98]$, assuming the tissues can be completely flattened. As both intervals do not overlap, we can distinguish pressure from touch and – since the interval for touch is limited to the left by a value of 93 DU – hover effects are also eliminated (cf. Figure 5.4).

A drawback of the depth camera approach is, however, that the scene has to be static. Whenever the user moves

an object, the depth values change. The extent to which this drawback affects the user study is listed in section 5.4—“Prototype Limitations”.

Recognition of Stationary and Non-Stationary Movement

As regards detection of planar movement and rotation of (everyday) objects, we decided to add a second technology to our prototype, developed by Kaltenbrunner [2009]. His *reactIVision vision engine* was originally developed for bringing physical interaction to multi-touch tables.

Paper fiducials are attached to the bottom of each object and are captured by a standard VGA camera mounted beneath a transparent table surface. The captured image is thresholded to black and white so that the fiducial pattern can be extracted. By analyzing the current orientation of the fiducial in the picture, information about position and rotation angle are retrieved. By putting a marker on a bottle cap, for instance, non-stationary planar movement can be recognized when the bottle is moved, as well as stationary movement is detected when the cap is screwed. Besides, each fiducial pattern is assigned an identification number. Figure 5.5 shows an example of a typical marker.

Paper fiducials feature a specific pattern of which position and angle can be retrieved using computer vision algorithms.

For the prototype, we attached a Microsoft VX-700 VGA camera to the mounting close to the Kinect such that both cameras shared nearly the same perspective (cf. Figure 5.1, right). Using the Kinect for marker detection was not possible since it was already occupied by the Kinect framework and not compatible with the *reactIVision* vision engine.

For fiducial detection, we added a web camera to the mounting.

Using paper-based fiducials brings great flexibility. They are freely scalable and can be easily attached to any object, e.g., by using double-sided adhesive tape. Besides, they do not alter physical affordances of an object unlike spheric markers used for the Vicon Motion Capture system³. As described, *reactIVision* fiducials feature a unique id. We

³www.vicon.com



Figure 5.5: A reactIVision fiducial. The specific pattern of the paper marker ($id = 0$, $angle = 30^\circ$) is recognized by the reactIVision vision engine with the aid of a web camera.

Paper fiducials were also used for recognizing user-initiated coupling and uncoupling of an object from the IUI.

exploited this idiosyncrasy for user-initiated activation and deactivation of an object for mode of appropriation via pointing or hiding. A small green marker with an odd id was put to the side of an object. Whenever the user pointed the object at the target system put behind the aluminum frame, the VGA camera recognized the marker and its id. For deactivation by hiding to function, each object was equipped with a second marker of white color representing an even id, put on top of the object. When the object was hidden, the marker disappeared from the camera image and the artifact was uncoupled from the IUI. The white fiducial was also used for obtaining object position and rotation angle. Figure 5.6 exemplifies the attachment of a green marker to a bottle a user is currently activating.

5.2.2 Linking Objects to Gestures and Actions

So far, we know how to turn passive objects into active controllers. Yet, what remains is how to establish a connection between object, gesture, and action to be executed when a user manipulates a part of an object.

The prototype architecture consists of three major components.

Figure 5.7 shows a detailed architecture of the prototype including the communication between its components, named *Kinect Controller* (A), *Fiducial Controller* (B), and *Event Processor* (C). Basically, the Event Processor is the core of the prototype and is responsible for forwarding events to the application to be controlled. The Kinect Controller is responsible for touch- and pressure detection



Figure 5.6: Object activation via pointing. The user points the bottle at the TV. Meanwhile, the small web camera (top) captures the green marker attached to the bottle. After three seconds of pointing, the bottle is activated.

and links each touch point to a method from the applications connected to the Event Processor. The Fiducial Controller manages object activation and deactivation plus forwards object position and angle to the Event Processor. So, logically, Kinect Controller and Fiducial Controller are responsible for recognizing input, whereas the Event Processor receives the input and processes it. Both Kinect Controller and Fiducial Controller communicate to the Event Processor via IPC (Inter Process Communication), being independent components. We will explain the functionality of each component subsequently.

A: Kinect Controller

Figure 5.8 shows a screenshot of the Kinect Controller Programming Interface (KCPI). It allows to graphically attribute touch- and pressure-sensitive regions to object surfaces. Imagine a stapler on a table. Using the KCPI, one first draws an object layer (green) around the object for

Touch detection for object surfaces is defined graphically.

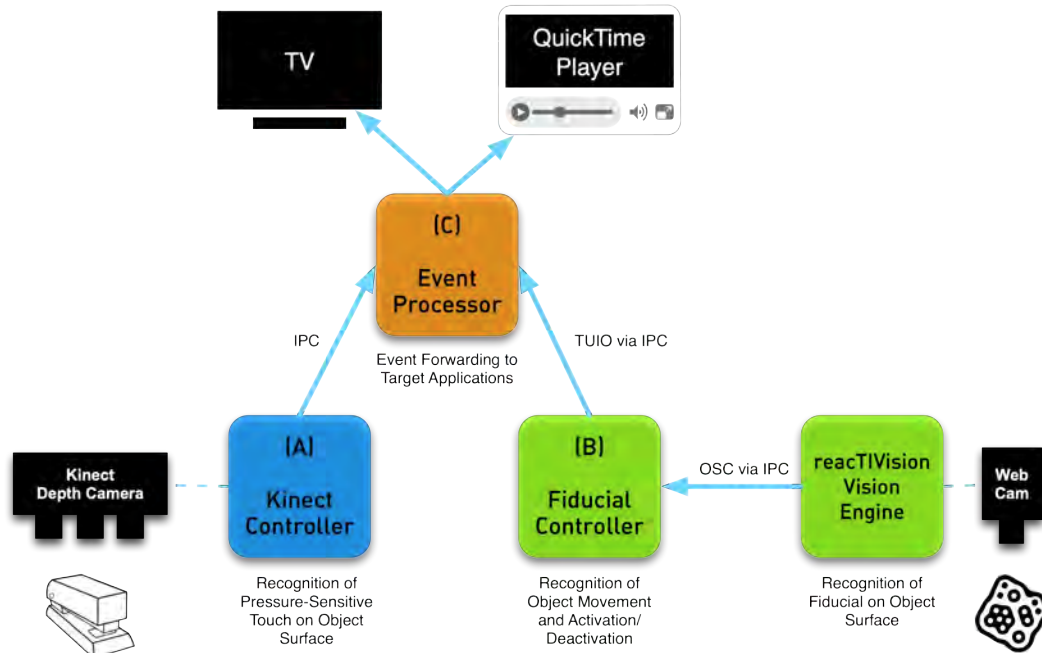


Figure 5.7: Prototype Architecture. The prototype consists of three components (A, B, C). The Kinect Controller is responsible for touch detection and forwards detected events to the Event Processor. Similarly, the Fiducial Controller is responsible for detection of fiducials put on objects. Whenever an object is moved, rotated, or (de)activated, the controller informs the Event Processor which is finally responsible for calling the methods of the target applications associated with the event.

determining the visual object scope. By drawing so-called touch zone layers (brown) on the object layer, touch- and pressure-sensitive regions are defined. For adding push detection to the stapler, object layer and touch layer would look as demonstrated in Figure 5.9. Of course, other parts on top of the stapler could be made responsive to touch, but those have no obvious push-affordance.

For each touch zone,
a depth range is
defined.

For each touch zone, we can define a depth range, as explained in 5.2.1—“Touch Detection on Arbitrary Surfaces”. Whenever the depth value of a touch zone changes to a value within the interval, a touch is recognized. Each touch layer can be linked to one method the Event Processor will call whenever a touch or push gesture is detected. Taking a music application into consideration, the recently created touch zone could be linked to a method called `play()`. When the stapler is pushed, music starts playing.

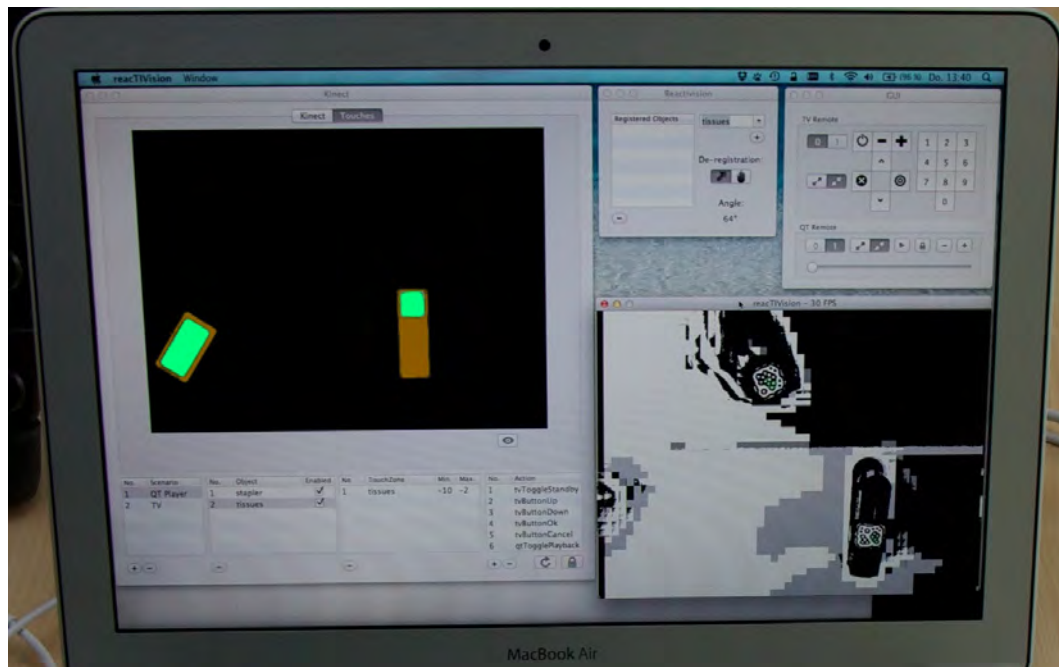


Figure 5.8: Principal investigator’s control panel. Left: Using the Kinect Controller Programming Interface (KCPI), the investigator can graphically define the object trace (brown shape) and which parts of the object surface should respond to touch/pressure (green shape). Each touch layer can be assigned a method from the “Action” table. Bottom: The reactIVision vision engine captures the fiducials put on objects. Top (right): Using the Fiducial Controller (left), the investigator can monitor registered objects. In case the investigator needs to interfere, he can control both the TV and QuickTime Player with a virtual remote control (right).

B: Fiducial Controller

The reactIVision vision engine captures the workspace beneath the VGA camera. For each paper fiducial recognized, it sends data, such as marker position, angle, and id to the Fiducial Controller using the Open Sound Control (OSC) protocol [Wright, 2005], a universal protocol akin to the MIDI⁴ protocol.

The reactIVision vision engine sends OSC messages to the Fiducial Controller.

⁴www.midi.org

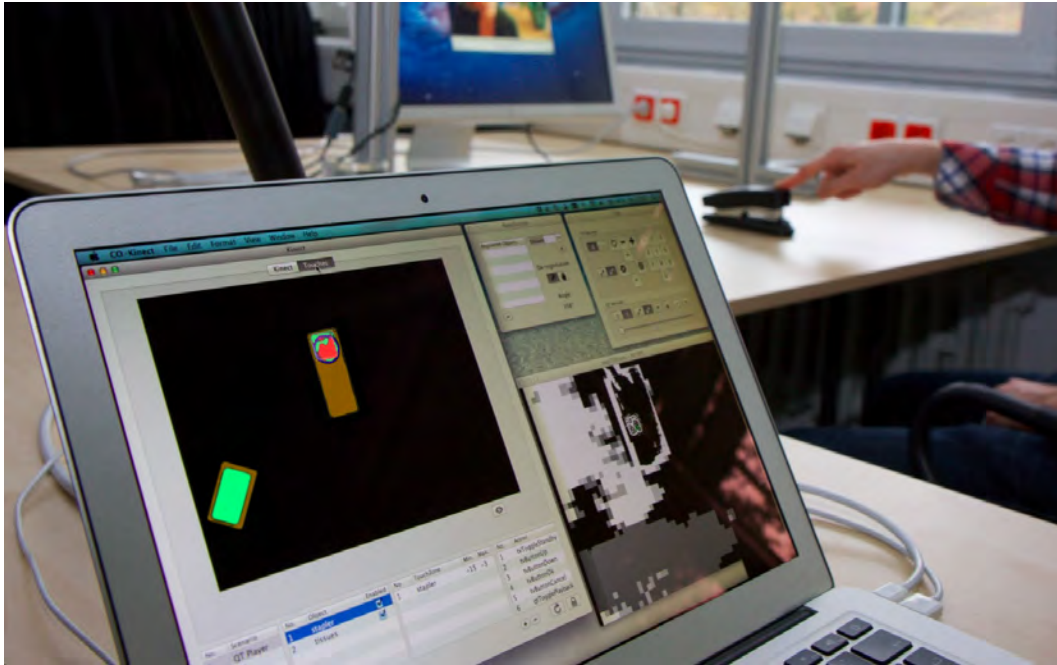


Figure 5.9: Touch detection for a stapler. The Kinect Controller (foreground) detects a touch event (red spot inside a blue circle) while the user (background) is pushing the stapler.

The Fiducial Controller wraps each OSC message received from the reactIVision vision engine into a TUIO-compliant message.

For reasons of compatibility, the Fiducial Controller wraps each OSC message received from the reactIVision vision engine into TUIO-compliant messages. “TUIO is an open framework that defines a common protocol and API for tangible multitouch surfaces. The TUIO protocol allows the transmission of an abstract description of interactive surfaces, including touch events and tangible object states [...]. Technically, TUIO is based on Open Sound Control – an emerging standard for interactive environments not only limited to musical instrument control – and can be therefore easily implemented on any platform that supports OSC.”⁵ This way, all fiducial-specific properties are directly accessible.

The Fiducial Controller communicates directly with the Event Processor.

Whenever an object is moved or turned, the Fiducial Controller sends the position coordinates and the angle to the Event Processor via IPC. Originally, the Fiducial Controller was integrated in the Event Processor. Yet, due to performance reasons we were obliged to outsource it.

⁵www.tuio.org

Besides, the Fiducial Component is also responsible for handling activation and deactivation of objects for appropriation. Whenever the user holds the green marker beneath the camera for at least three seconds, activation and deactivation are toggled. Deactivation by hiding is achieved when the white marker has vanished for at least three seconds from the camera. For each activation and deactivation process, the Fiducial Controller informs the Event Processor via IPC. Figure 5.8 shows the reacTIVision vision engine and a screenshot of the Fiducial Controller.

The Fiducial Controller is responsible for handling object activation and deactivation.

C: Event Processor

The Event Processor processes all messages sent by Kinect Controller and Fiducial Controller via IPC. Before any incoming event is processed, the Event Processor checks whether the addressed object has been activated. Incoming events of deactivated objects are ignored. Concerning push and touch events, the name of the method to be called is wrapped in the message and will be executed instantly. For objects connected to the Fiducial Controller, position and angle values are redirected to the target application, which is finally responsible for the interpretation of the values.

Incoming events from deactivated objects are completely ignored.

In the scope of this user study, we implemented a TV application and a QuickTime Player controller. The latter enables users to control [Apple QuickTime Player](http://www.apple.com/quicktime)⁶ via everyday objects, whereas the first one enables participants to operate a television set with quotidian objects. Although not flexible, we integrated both applications into the Event Processor, which is fully sufficient for testing.

We implemented two applications that could be controlled with everyday objects.

The prototype software presented has been developed for OS X Lion. IPC over sockets was done with [ThoMoNetworking](http://hci.rwth-aachen.de/thomonet)⁷.

⁶www.apple.com/quicktime

⁷hci.rwth-aachen.de/thomonet

5.2.3 Applications

Revisiting Figure 5.7, the Event Processor is connected to two target applications, specifically designed for the subsequent user study. Both applications represented scenarios that have been inspired by the WOZ study.

We simulated a TV interface and implemented a movie player controller for testing everyday object interaction.

Target I simulated a television interface, including various TV channels and an OSD (On-Screen Display). Referring to the small isolated GUI scenarios tested in the previous study, we decided to test a real application that combined several UI widgets, such as buttons and a slider. Therefore, we implemented a controller that communicated to a standard video player application (Target II). The functionality of both target applications is described briefly.

Television Application

The TV application featured true video material recorded from 15 channels.

The television application was patterned on standard modern television interfaces that feature a graphical channel list (cf. Figure 5.10, top) and an OSD menu (cf. Figure 5.10, bottom). Although this application needed to be run on a computer, it looked like a natural TV: The television image was full screen without any window borders that might have reminded of a WIMP interface. The virtual TV had 15 channels in total, including four of them in HDTV (High Definition Television) quality. For each channel, we recorded approximately ten minutes of true television video material. Playback was looped, i.e., after ten minutes the video stream of each simulated channel started to play again from the beginning which was never experienced by the testers during the study.

The TV featured an OSD menu.

As regards the menu, it contained five selectable options, i.e., “Channels”, “Timer”, “Guide”, “Brightness”, and “Turn Off”. Selecting the first entry faded in the channel list, “Brightness” allowed to adjust the brightness of the television image, and “Turn off” enabled the user to switch the TV off via the menu. The remaining two options were not implemented since they were not planned to be used for the study.

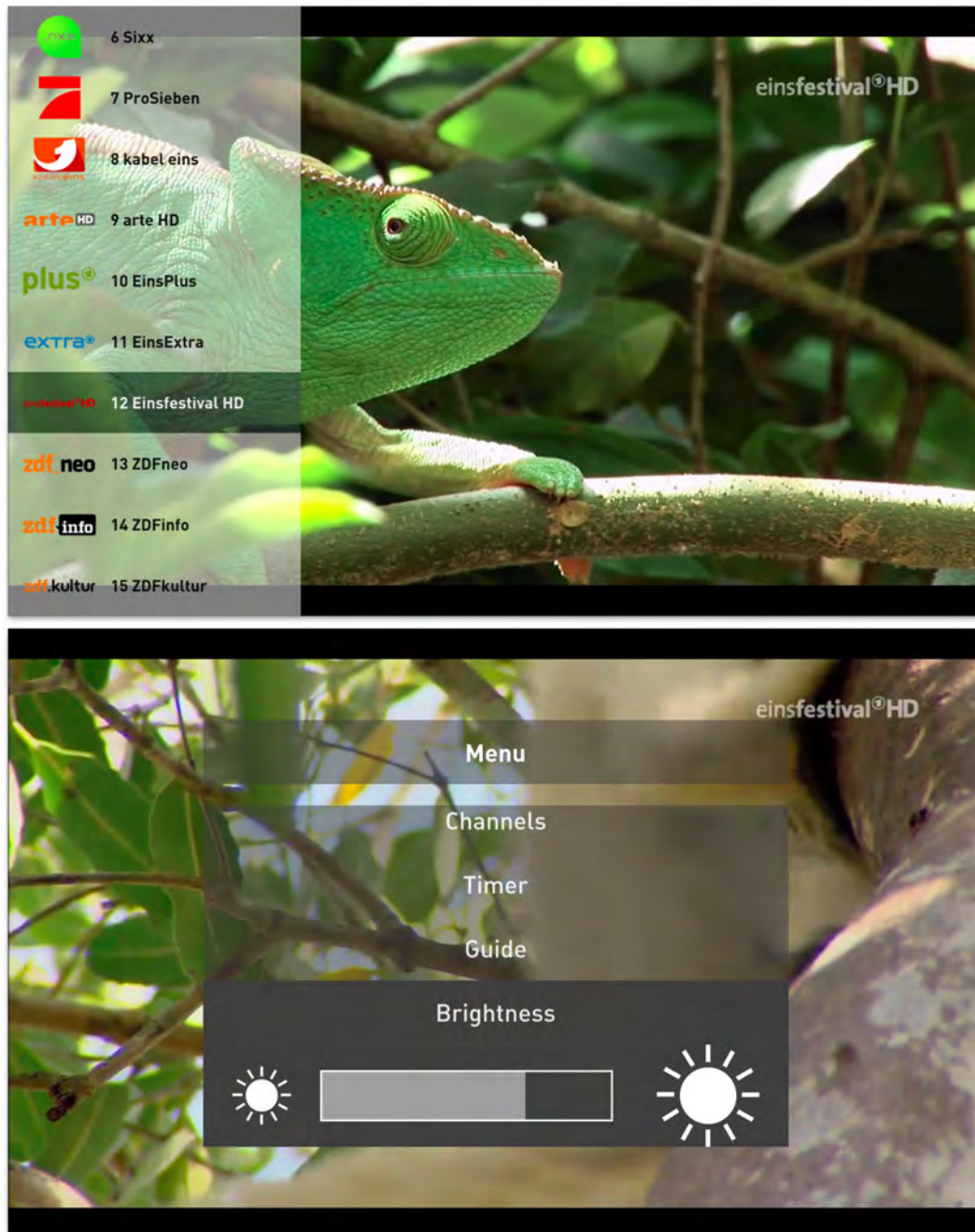


Figure 5.10: TV channel list and menu. The channel list (top) featured 15 channels that could be browsed up and down by pushing the tissues. The OSD menu (bottom) consisted of several sub menus, such as brightness control.

The principal investigator had access to a virtual remote control; testers were only allowed to use everyday objects.

For the principal investigator, the television application featured basic navigation as known from standard remote controls. This enabled him to interfere whenever needed. The virtual remote control is shown in Figure 5.8. Please note that this virtual control was not accessible by the testers since we wanted them to control the TV interface using everyday objects. Besides, a channel could be accessed by navigating the channel list up and down and by finally confirming the selected channel. Menu navigation is done the same way.

QuickTime Player Application

The QuickTime Player controller featured standard playback controls.

Concerning Target II, we chose QuickTime Player application, version 7.6.6. Using the [Apple ScriptingBridge framework](#)⁸, we implemented a small controller application which forwarded incoming events from the Event Processor to the original QuickTime Player application. Standard commands were play/pause, rewinding and forwarding the video or adjusting the volume using a horizontal slider, and full screen playback. Figure 5.11 shows a screenshot of the QuickTime Player 7.6.6 GUI.

5.3 Qualitative User Study

We conducted a think-aloud user study.

Based on the goals defined in section 5.1—“Goals”, we conducted a qualitative user study testing everyday object interaction for two scenarios based on the applications presented: (A) controlling a TV, and (B) controlling a computer-based movie player. The testers had to accomplish a sequence of tasks with a set of everyday objects at hand. While the study was in progress, the testers were asked to comment on what they were doing (think aloud). Each session was recorded on video for qualitative analysis.

⁸<http://developer.apple.com/library/mac/#documentation/Cocoa/Conceptual/ScriptingBridgeConcepts>



Figure 5.11: QuickTime Player GUI. The users were asked to control QuickTime Player 7 using everyday objects. The movie player featured playback/pause, volume control and fast forwarding/rewinding via a horizontal slider, and a full screen mode.

We will first present the physical setup of this study, followed by a presentation of the study procedure, alongside with a short demographic analysis on the study participants and study limitations, before evaluating results.

5.3.1 Setup

The physical setup of the study room is depicted in Figure 5.12. The prototype consisting of the aluminum mounting for the Kinect and the web camera was put on a table, sized 115x74 cm² (width x depth). Both cameras shared nearly the same viewpoint at bird's eye view on the desk surface. Approximately 40 cm behind the mounting, we put a 23-inch Apple Cinema HD monitor which served

The table surface beneath the aluminum mounting served as user's workspace.



Figure 5.12: Study setup. The user (sitting on the chair) was meant to manipulate objects on the table beneath the aluminum mounting. The monitor behind the mounting served as display for the scenario applications. Object manipulation was captured by camera (fixed on a tripod). The principal investigator sat behind the laptop which served as control panel.

as display for the target applications. The monitor and the cameras were connected to a mid-2011 MacBook Air. The table surface beneath the cameras and in front of the monitor served as interaction area, sized approximately $58 \times 44 \text{ cm}^2$ (width x height), where the user was meant to manipulate everyday objects.

For each scenario, the user had four objects available for manipulation.

For each scenario, the users had four objects at their disposal which were initially put beside the workspace on the table. For the TV application, the participants could interact with (1) a pack of pocket tissues, (2) a cup, (3) a wallet, and (4) a glass bottle. QuickTime Player could be controlled with (1) a stapler, (2) a glue stick, and – as given in the TV scenario – (3) a pack of pocket tissues and (4) a glass bottle (cf. Figure 5.13).



Figure 5.13: Everyday objects available for the user study. Cup, wallet, bottle, and pocket tissues were available for the TV scenario. The latter two items were also available for the movie player. Besides, the stapler and the glue stick were used for this scenario. Applicable gestures for each object are given in Figure 5.14.

Next to the table, we set up a digital video camera fixed to a tripod. It captured both the user interacting in the workspace and the monitor for later analysis.

5.3.2 Participants

In total, 16 participants, aged 23–61 ($M = 29.50$, $SD = 12.01$, five females), took part in the study. Nine users were computer science students; two participants studied medicine. Besides students, a teacher, a major, a system consultant and a cosmetician participated in the study. Ten users had also participated in the WOZ study. However, since the WOZ study took part three months before, they did not remember their executed gestures from the latest study, so that their “prior experience” did not influence the

In total, 16 users participated in the study.

results of this study. On average, the testers stated to watch occasionally TV but agreed to watch frequently movies with a computer-based video player.

5.3.3 Tasks

In order to simulate the interplay of normal object usage and appropriated use, tasks were mixed with interrupts.

For each scenario, the users had to accomplish a sequence of given tasks. The only tools allowed were the everyday objects at hand. Alternating with the tasks related to the target application, the users were also asked to complete object-related tasks, so-called interrupts. This mixture of tasks was to represent the interplay of appropriated object use and normal object usage. In the movie player scenario, for instance, the participants were demanded to staple some sheets of paper right after they had played back the video. Table 5.1 and Table 5.2 list the sequence of tasks for both scenarios including the interrupts and possible side effects that could occur when an interrupt was executed without having deactivated the object beforehand.

The sequence of tasks and interrupts was unaltered.

The sequence of tasks for both scenarios was identical for each user since some tasks required previous tasks to be accomplished first. Turning the volume louder, for instance, presumed that the TV had been switched on previously.

Task/ Interrupt	Goal	Solution Possible Side Effects
T1	Power the TV on.	(a) Push the up-/down-“button” on the tissues. (b) Turn the cup handle to approximately two to three ‘o clock.
I1	(a) Put the tissues into your pocket. (b) Imitate drinking from the cup.	(a) Pushing the tissues while grabbing them scrolls the channel list or fades in the menu. (b) Turning the handle for grabbing/putting the cup back could switch the TV off.
T2	Go to channel “einsfestival HD”.	Push the down-“button” on the tissues twelve times (opens the channel list automatically and selects the channel) and push the center on the tissues to confirm the selected channel.
I2	Put the tissues into your pocket.	Pushing the tissues while grabbing them scrolls the channel list or fades in the menu.
T3	Increase the volume to the maximum.	Turn the bottle or its cap clockwise.
I3	Open the bottle and pour some water into a glass.	Screwing the cap off or moving the bottle changes the volume.
T4	Go to channel “Das Erste HD”.	Push the “up”-button on the tissues twelve times (opens the channel list automatically and selects the channel) and push the center on the tissues to confirm the selected channel.
I4	Open the wallet.	Pushing the wallet while grabbing it fades out the channel list, menu, or volume indicator (if visible).
T5	Decrease the brightness to some value.	Push the center on the tissues to open the menu, push the down-“button” to select the “Brightness” sub menu, push the center to enter the menu, and turn the bottle or its cap counter-clockwise.
T6	Power the TV off.	(a) Turn the cup handle to approximately twelve ‘o clock position. (b) <i>if still in brightness mode</i> : Push the up- or down-“button” on the tissues to escape the sub menu, push the down-“button” on the tissues to select “Turn off”, and push the center to confirm. (c) Push the center on the tissues to open the menu, push the down-“button” four times to select “Turn off”, and push center to confirm.

Table 5.1: Tasks and solutions for the TV scenario. The table presents the demanded tasks (T) alongside with a description concerning how an object had to be manipulated to achieve the goal. Alternating interrupts (I) were added to simulate normal object usage. Letters in brackets indicate alternatives.

Task/ Interrupt	Goal	Solution Possible Side Effects
T1	Play the video.	(a) Push the stapler. (b) <i>Indirect Solution:</i> Push the tissues to enter full screen mode. QuickTime Player will automatically start playback.
I1	(a) Staple some sheets of paper. (b) Put the tissues into your pocket.	(a) Pushing the stapler could toggle playback mode. (b) Pushing the tissues while grabbing them toggles full screen mode.
T2	Decrease the volume to some level.	Turn the bottle or its cap counter-clockwise.
I2	Remove the bottle cap.	Screwing the cap off or moving the bottle changes the volume.
T3	Pause video playback.	Push the stapler.
T4	Change the video time position to approximately 18 minutes.	Move the glue stick horizontally to the right until the time position is correct.
I3	Put some glue on a sheet of paper.	Grabbing the stick changes the currently set time position of the video.
T5	Enter full screen mode.	Push the tissues.
I4	Put the tissues into your pocket.	Pushing the tissues while grabbing them toggles full screen mode.
T6	Play the video.	Push the stapler.
T7	Exit full screen mode.	Push the tissues.

Table 5.2: Tasks and solutions for the QuickTime Player scenario. The table presents the demanded tasks (T) alongside with a description concerning how an object had to be manipulated to achieve the goal. Alternating interrupts (I) were added to simulate normal object usage. Letters in brackets indicate alternatives.

5.3.4 Procedure

The user was free to try all objects out that were available per scenario.

First, the user was introduced to the capabilities of the prototype by telling her how to activate an object (pointing), how to deactivate it (pointing vs. hiding), and which types of gesture were supported (touch/push, move, turn). Whenever the user thought activation/deactivation was necessary, she was free to do so. For a given task, the user could principally pick any object she thought that would fit best for accomplishing the task. However, mappings be-

tween object, gesture and action were predefined inspired by the mental models from the WOZ study. End user programming was not tested. The objects had no visual cues, neither for the activation state nor for the regions that were interactive. All objects were equipped with markers for activation, deactivation, and for detection of movement. However, not each object did react to movement or rotation, so the markers were no indicators for the user that a certain gesture type was supported by the object or not. Similarly, touch- or pressure-sensitive areas of an object were only visible to the principal investigator via the KCPI. Figure 5.14 illustrates which gestures were recognized for an object and which actions were linked to them. Having picked an everyday object, the user could try out different gestures and look for the outcome on the screen. At any time, the user was allowed to add or pick a different object (out of the four available).

As a consequence, it could happen that users previously explored how an object was supposed to be appropriated for a task before this task was due. For example, if somebody used the bottle because she thought that moving it would switch the channel, she could see that the volume indicator was faded in and its level was slightly changed (because turning is accompanied with slight lateral movement). From now on, the user knew that volume could be controlled with the bottle. As said before, randomization of tasks was not possible, because a task was dependent on the previous one. Consequently, objective analysis of intuition of the predefined mappings, e.g., by counting the number of items that were tried out before a task was solved, could not be done. Thus, we focused on qualitative user feedback as regards intuition of predefined mappings (cf. 5.5.1—“Users’ Performance on Tasks”).

Whenever the user felt unable to solve the task (e.g., after several tries of objects and gestures), she could abort it. After a scenario was completed, the user was asked to reveal her interpretation of the objects as experienced in the study. Having completed both scenarios, the user was further asked to fill out a questionnaire regarding interaction experience for each scenario as well as general questions.

Due to trial and error strategy, a user could anticipate a predefined mapping before the task was due.

In the end, the users were asked to fill out a questionnaire.



Figure 5.14: Predefined gestures for appropriated objects. (A) – (D) show the object-gesture mappings for the TV scenario. The cup handle, for instance, was meant to be turned to twelve 'o clock position for turning the TV off; turning the handle to two/three 'o clock position switched the TV on. (E) – (H) illustrate the mappings for movie player control. Red dashed shapes separate touch-sensitive regions.

5.4 Prototype Limitations

Although the depth camera approach is a powerful means for touch and pressure augmentation on everyday objects, it has a major drawback: Objects are not allowed to be moved during interaction as this alters the depth values and – consequently – interferes with the accuracy of touch recognition. Basically, one could argue that adding object tracking paired with dynamic re-calculation of the depth values is a solution to this problem. Although this is theoretically correct, it has no practical value since the user occludes the object by hand while moving it. Hence, this effect distorts depth calculation since not only the naked object is scanned, but also the hand put on it. Therefore, the user was told to inform the principal investigator whenever she was about to touch or push an object after she had moved it. Then, the principal investigator would re-drag the object layer to the new object position and prompt the Kinect to recalculate the depth values. Although this procedure interrupted the fluidity of interaction, the participants did not feel negatively influenced in their interaction. Regarding planar movement, users could freely turn or replace objects as long as they did not exceed the limits of the workspace.

Object movement
distorts depth
calculation.

A further limitation made by the Kinect is the size of touch-sensitive layers. Initially, based on the patterns observed in the WOZ study, we planned to add a ball pen or a mobile phone as button-equipped objects to the object repertoire. Unfortunately, the resolution of the Kinect was too low to obtain satisfying and reliable touch events on small surfaces. Consequently, small objects could not be manipulated and were not tested in this study.

Unfortunately, the
resolution of the
Kinect was too low
for detecting small
touches.

5.5 Findings

From the video analysis on interaction with everyday objects, we obtained a wealth of information. While analyzing the video recordings and the questionnaire, we were basically interested in two aspects:

- **Solution of Tasks.** Which tasks worked well and which tasks caused problems? What were the reasons for interaction flaws?
- **Interplay of Normal and Appropriated Object Use**
How did people handle activation and deactivation of objects and to what extent did side effects occur/-play a role?

5.5.1 Users' Performance on Tasks

As regards the users' task performance for both scenarios, we identified three sources producing interaction flaws:

1. **Technical Problems.** As touch- and pressure detection required the object to be in a fixed position, sometimes gestures were either misinterpreted or not recognized when their execution was accompanied by unintended slight movement of the artifact. Sometimes, delayed gesture recognition and system feedback was observed.
2. **Mental Models.** If the user had a mental model that differed from the predefined object-gesture-action mappings, she needed to try out different mappings until it matched the predefined one, delaying interaction. In the worst case, a matching mapping could not be found. The task could not be solved.
3. **Object Activation/Deactivation.** Despite having applied one of the predefined gestures, the expected reaction by the system did not happen when the object was not activated. Forgotten deactivation could lead to side effects when dealing with interrupt tasks.

Technical problems could be solved by means of system re-calibration.

In terms of technical problems, the prototype could be easily re-calibrated or restarted during the study so that the impact of this error source was minimal. For future studies, these flaws can be eliminated with more sophisticated technology (cf. 6.2—"Future Work"). Therefore, we do not focus on this source of error.

We concentrated on interaction flaws directly dealing with the user, referring to sources (2) and (3). In fact, most interaction flaws were based on a mismatch between the user's anticipated object-gesture-action mappings and the mappings that were actually implemented for the prototype. The second-most frequent source of error identified was due to object (de)activation.

Mismatches between a user's mental model and predefined gestures were observed.

Firstly, we will analyze the user's task performance based on gesture mappings. Secondly, we will investigate activation and deactivation mode of objects. Finally, we will summarize the outcome of the study and draw a conclusion on the interpretation of the results.

Table 5.3 shows the results for the users' success in completion of the demanded tasks for each scenario by contrasting accomplished tasks with aborted tasks. Comparing the overall task success, the movie player won over the television scenario. Where appropriate, results of the feedback form will be mentioned. If not otherwise indicated, average rating based on a five-point Likert scale is considered, 1.0 meaning "strongly disagree", 5.0 representing "strongly agree". Questions are referenced by Qn , where n is the number of the question as listed in the questionnaire. The feedback form is listed in B—"Appendix for the Interactive Prototype Study".

Regarding general task success, the movie player scenario won over the TV scenario.

Task Performance for the Movie Player Scenario

As regards QuickTime Player, pushing the stapler for toggling playback and pause (Tasks 1/6, 3) as well as turning the bottle for controlling the volume (Task 2) were accomplished by all participants. For these tasks, the predefined object-gesture mappings were considered intuitive, supported by the following comments:

Playback/pause and volume control were handled by all participants.

- "It [the bottle] worked like any volume controller. Good mapping!"
- "It seems that turning the bottle alters the volume. It makes sense."

Television Scenario		Movie Player Scenario	
Task No.	Solved by [Users]	Task No.	Solved by [Users]
1	15	1, 6	16
2	10	2	16
3	16	3	16
4	10	4	16
5	4	5	13
6	15	7	13

Table 5.3: Task success for both scenarios. In general, users were more successful at controlling the movie player than the TV. As regards the movie player, only entering/exiting fullscreen was problematic for three users. For the TV scenario, altering the brightness was most challenging, whereas all users ($N = 16$) managed volume control.

- “It [the bottle] works quite well for the volume.”
- “Very intuitive for pushing [the stapler] because of the haptic feedback.”
- “That [the stapler] can be pushed like a button. It’s good for playback.”

Changing the volume with the bottle was perceived easiest among all movie player scenarios.

Intuitiveness was also confirmed by the results of the feedback form: Asked whether the mapping between object, gesture, and executed action were as expected, volume change by turning the bottle got the highest score (Q8b: $M = 4.56$, $SD = 1.17$), followed by playback and pause by pushing the stapler (Q8a: $M = 4.19$, $SD = 1.17$). The perceived intuitiveness of the predefined volume mapping correlated with the ease of solving the task: On average, the users found this task easy (Q6b: $M = 4.06$, $SD = 1.00$). In fact, it was perceived easiest of all scenario-related tasks.

The mapping for the play/pause button was adopted from the check box scenario (WOZ study).

For the play/pause button for the video player, we adopted the most frequently observed pattern from the check box scenario (WOZ study), as both controls are toggles for two different states. As this predefined mapping also worked very well (cf. above), we see our observations from the WOZ study confirmed.

All participants managed to control the playback position with the QuickTime Player (Task 4). And yet, interaction flaws were observed, confirmed by the results of the feedback form. The users did not fully agree that the task was easy to solve (Q6c: $M = 3.81$, $SD = 0.91$). Inspired by the patterns identified for the slider scenario (WOZ study), we did a 1:1 mapping: The participants were meant to move the standing glue stick from the left table side to about $\frac{3}{4}$ to the right of the interaction space, thus they had to *slide* the stick (cf. Figure 5.15). The virtual slider moved likewise. In total, six participants took the glue stick straight away – they associated slider control with it. Nonetheless, only three participants moved the stick horizontally right from the beginning. All other participants first tried to turn the stick clockwise – they thought this would make the timeline slider move to the right and – as in standard movie players – consequently forward the video. We could identify two reasons why 13 out of 16 participants expected turning for slider control.

To our surprise, 13 out of 16 participants expected the slider to be controlled by turning.

First, the glue stick was influenced by the bottle. Both objects share a basic geometric form, i.e., they are tall and cylindrical. Since all participants experienced turning the bottle for continuous volume control before, the participants expected the similarly shaped glue stick to work the same way - by turning. One student commented: *"The glue stick was influenced by the bottle I used before. At the beginning, I turned the stick."* Another participant even expected that turning any cylinder should always result in the same action: *"The bottle and the glue stick have a similar shape, but they had different functions."* People were fixated to the shape of an object and expected a consistent gestural use and effect. Thus, the "form follows function"-principle [Dix et al., 2003] applied here: A cylindrical object affords the action of turning to a stronger extent than sliding it for movement. Originally, in the pilot study we used a magnet instead of a glue stick for slider control. The magnet was meant to move like the glue stick for controlling the slider which was inspired by the most frequently observed pattern in the WOZ study. However, as the magnet was relatively flat and therefore farther away from the web camera than the stick, we had to put a large marker on the magnet for the camera still to be able to recognize the fiducial pattern. This influenced interaction to the

Cylindric objects afford turning.



Figure 5.15: Glue stick in action. A user is sliding the glue stick to the right for dragging the timeline slider to the same direction in order to forward the video.

worse, since the marker was bigger than the magnet itself. Therefore, we decided for the second most frequently observed pattern, with a slight modification: Instead of using a standing whiteboard marker, we chose the glue stick as it had better footing. Despite turning the glue stick, all participants agreed that the 1:1 mapping of horizontal movement made perfect sense and felt natural, once the testers discovered that the glue stick had to be moved like a slider. This was also reflected in the evaluation of whether the predefined mapping was suitable for solving the task, since the testers basically agreed to it (Q8c: $M = 4.00$, $SD = 0.89$). Interestingly, one user commented: *"I expect to turn the glue stick to slide through the video."*. Thus, although he used "slide", he anticipated turning.

As a consequence to the first argument, the second reason why people applied turning was, because it actually resulted in a position change of the playback position. The testers interpreted this as confirmation of being on the right track. In fact, turning an object results in a slight

movement to all directions, thus also to the left and to the right. Consequently, this slight horizontal movement was interpreted by the system as sliding, hence resulting in a change of the playback position. This interaction flaw therefore led to wrong interpretation by the users (cf. 5.5.2—“Users’ Interpretations” and shows the importance of testing everyday object interaction with real time feedback. In cases where movement gestures are involved, one must carefully filter out side effects, such as peripheral motion. Otherwise, it can lead to misinterpretation and result in the user’s withdrawal.

Turning was accompanied by unintended lateral movement which was misinterpreted by the system.

Therefore, the two reasons mentioned strengthened the testers in their interpretation that cylindrical objects had to be turned. Of course, these objects *can* do just that. They can be used for sliding just as well though, as the WOZ study confirmed.

Pushing the pocket tissues for toggling fullscreen playback of the video (Tasks 5, 7) was least intuitive for the testers. In total, three testers did not solve this task. The participants did neither agree nor disagree that the mapping was suitable (Q8d, $M = 3.44$, $SD = 1.15$). Actually, this mapping was perceived most unintuitive for this scenario. Some participants did not associate pushing with the pack of tissues. Often, people first tried to move, rotate, or squeeze it. For example, tester C. L. commented: *“The tissues are for squeezing. They have no push-affordance to me.”*. However, user J. K. praised the intuition of pushing the tissues: *“The first thing I would do to the tissues is push them.”*, which is exactly what he did for both the movie player and the television scenario. As regards the WOZ study, pushing a pack of pocket tissues was observed as a pattern only for the television scenario. Yet, pushing the tissues was scattered across other scenarios, but not as a common pattern. We therefore believe that tissues – despite the haptic feedback they provide when pushed – are not a useful solution in this context, i.e., as substitute for a button. Not surprisingly, for the testers there was no clear connection between a pack of tissues and the linked action of entering fullscreen mode. Paired with no clear push affordance, the pack of tissues is not a feasible solution for alternatively representing a fullscreen button. Interestingly, the haptic feedback received for pushing a

Appropriating the pocket tissues as push button for fullscreen playback was considered unintuitive.

pack of tissues is not consistent: A full pack (as used in this study) usually consists of ten tissues, providing a strong response to push. The fewer tissues remain in the pack, the less haptic feedback is provided. Therefore, if we had used a nearly empty pack of tissues, the push affordance would probably have been even less noticeable.

People probably match the strongest push-affordance to the main function of a GUI.

Comparing the tissues with the stapler, we also saw no obvious connection to the action of playback. In this context, J. K. stated: *"When I push a stapler, I assign the action of stapling. That does not match with play/pause. I do not want to staple the video."* Despite the missing link, the stapler worked well in principle for playback. In fact, eight persons took the stapler right from the beginning and pushed it when asked to play the video. We believe that people match the device with the strongest push-affordance to the main task of the application to be controlled – video playback in this case. Among all objects used in the movie player scenario, the stapler had the strongest affordance for being pushed. Even J. K., who did not want to use it for playback, said: *"The stapler is very intuitive for pushing because of the haptic feedback."* Verification of this assumption is considered in chapter 6.2—"Future Work".

Task Performance for the Television Scenario

Volume control for the TV was consistent with volume control for the movie player.

For reasons of consistency, volume control for the television scenario (Task 3) was identical to the video player scenario, as were the results for task completion: All participants solved this task. Taking the results of the feedback form into consideration, intuitiveness of controlling the volume by turning the bottle was rated slightly lower (Q3d: $M = 4.25$, $SD = 0.58$) than for the other scenario. A positive learning effect could be observed in seven cases for this task: As the sequence of scenarios was mixed, a tester either experienced volume control with the bottle first in the television scenario and then in the movie player setting, or vice versa. Seven persons, who did not directly find out that the bottle had to be rotated for volume change in the first scenario, did so straight away in the second scenario. In two cases, we cannot say whether a learning effect existed, as they noticed that turning was used again

for volume control in the second scenario due to prior experience when trying out objects for the tasks that preceded volume control. As known from Nielsen's ten usability heuristics, consistency results in good usability [Nielsen and Mack, 1994]. This was also commented by user M. B.: *"The volume control is consistent; and consistency is good."* Paired with the results from the QuickTime Player Scenario, we conclude that controlling the volume by appropriating a bottle works well. On average, this task was considered easiest among all tasks in this user study (Q1c: $M = 4.56$, $SD = 0.51$), hence even slightly easier than the volume task from the video player scenario. Figure 5.16 shows volume control with the bottle in action.

Although turning the bottle for volume control was not the top pattern observed in the TV scenario (WOZ study), it was a working mapping, though. The mobile phone (cf. 4.4.3—"A2: *Watching Television*") had technical limitations – its buttons were too small and thus it could not be used.

Channel switching using a pack of pocket tissues (Tasks 2, 4) was only accomplished by ten participants. We identified three reasons leading to interaction problems:

1. People tended to attribute only one function to an object. The TV could be turned on by both pushing the virtual up/down button on the tissues and by turning the cup handle to approximately two to three 'o clock. For most participants who used the tissues for turning the TV on (previous task), the tissues were already occupied with this function: *"Oh, this is for turning on."*, one user commented. These users were intuitively looking for another object when they were asked to switch channels. Hence, for them, tissues were only good for power management.
2. Similarly, some people tended to assign just one virtual button to an entire object, as long as there were no physical or visual cues that could hint at multiple virtual buttons linked to the object. In fact, the entire object represented one button for them. Hence, they primarily pushed the center of the tissues, which did not result in channel switching. However, the "one

Using the tissues as remote control replacement caused interaction problems.

Basically, people only attributed one function to an object.

In general, users did not subdivide a continuous surface into smaller parts.



Figure 5.16: Bottle in action. A user is turning the bottle clockwise in order to increase the volume of the TV. Visual feedback is given via the volume level indicator.

object has just one function"-mapping was not consistent with the observations made in the WOZ study, where people tended to attribute multiple functions to a single object, e.g., for imitating a remote control. Yet, mapping just one function to the pack of tissues as experienced in the movie player scenario felt easier to handle than having multiple functions mapped to the artifact.

3. For some users, channel switching was associated with movement, therefore they tried out several objects and moved them up and down on the table surface. Participant C. L. commented that *"moving objects did not have any reaction in this scenario [i.e., channel browsing]. It's not expected."* However, it must be noted that vertical object movement is impractical while watching TV, since it requires the user to sit at a table to put the objects on. Interestingly, movement gestures were not observed in the WOZ study for watching television. We believe that this is due

Often, channel switching was associated with movement.

to the visual style of the channel list, which was not available in the WOZ experiment. For this study, the channel list was automatically faded in for ten seconds when the TV was turned on. Therefore its layout was perceptible to the users even if they did not manage to switch channels.

The channel zapping task showed that the object shape played an important role as indicator for a potential function of an object. For example, J.K. stated: *"The tissues look like a remote control."*. Indeed, he intuitively subdivided the tissues into at least two different parts: the upper part for switching the channels in one direction, and the lower part for switching the channels in the opposite direction (cf. Figure 5.17). Although the majority of the testers solved the tissue tasks, it is not obvious that an object – which has no visual cues for virtual buttons – is considered to have multiple touch-sensitive regions (cf. above). Yet the shape of the object plays an important role when it resembles the device to be replaced. In this case, there is a chance that people assign multiple virtual buttons including the correct positioning of them to the everyday object (as proven by ten testers).

The shape of an object played an important role as regards appropriation.

Interestingly, although only ten people solved channel browsing successfully, all testers overall agreed that the predefined mapping was suitable (Q3c: $M = 4.00$, $SD = 0.89$). As regards the ease of task handling, the results matched the observations made during the study (Q1b: $M = 3.25$, $SD = 1.06$), i.e., it was perceived more difficult than controlling the volume, for instance.

Despite difficulties, using the tissues as remote control replacement was considered a suitable substitute.

Brightness control (Task 5) with everyday objects was perceived most difficult for the users. Only four testers were able to solve this task and only one of them applied the predefined gesture sequence straight away, i.e., without trying out other objects first. The feedback form showed that the users did not agree that this task was easy to solve (Q1d: $M = 2.19$, $SD = 1.42$). The difficulties originated in the fact that people were not looking for a menu for activating the brightness mode (as is usually the case in modern television sets). Having everyday objects at reach, they expected a direct link between object and brightness

People expected direct access to functions via everyday objects; modes were not expected.



Figure 5.17: Tissues in action. A user is pushing the lower third of the pack of tissues with his forefinger for navigating the channel list one step down.

control, hence not via a sub menu. User, C. L. commented: *"I imagined the bottle rotation could adjust the brightness like the volume, but it did not."* Participants expected turning, as that was considered similar to volume control. Turning the bottle did in fact control the brightness but only when brightness mode was activated via the main menu which had to be navigated with the tissues (cf. Figure 5.18). Since channel switching with tissues was similar to menu navigation and success was only medium (cf. above), performance for the brightness task was even worse. This task showed that direct access to a function via an appropriated object is expected and not navigation through a menu or by activating a mode.

As regards switching the TV on or off (Tasks 1, 6), 15 participants were able to handle the tasks. On average, most of the users agreed that those tasks were easy (Q1a:



Figure 5.18: Bottle in action. Similar to volume control, brightness was controlled by turning the bottle or its cap. However, prior to that, brightness mode had to be activated via the menu by using the tissues.

$M = 3.81$, $SD = 0.83$). Here, redundancy was involved, as power management could be achieved both with the cup and the tissues. In six cases, the cup was used for turning the TV on. For turning the TV off, it was used eleven times. Overall, the tissues were less popular: Nine (four) testers used the tissues for switching the television on (off). Despite higher popularity for the cup, the participants considered pushing the virtual up or down button on the tissues a more suitable mapping than turning the cup handle (Q3a: $M_{cup} = 2.75$, $SD_{cup} = 1.29$, Q3b: $M_{tissues} = 3.56$, $SD_{tissues} = 1.15$). We think that the tissues were regarded more intuitive since the gestures linked to them were based on remote controls of modern television sets: Pushing the channel up or down button usually wakes the TV up from standby. Yet, turning the TV off with the cup was

Pushing the tissues was considered a better mapping for turning the TV on than turning the cup handle.

much more popular than using the tissues. This was due to the following reason: As in the brightness task, turning the TV off with the tissues required the user to enter the main menu and navigate to the “Turn off” entry. This procedure was too complicated. As the brightness task was preceding, two of the four users who succeeded in that task used the tissues for turning the TV off. Once they were navigating the menu anyway, and the “Turn Off” entry was visible directly underneath the brightness sub menu, they quickly accomplished the task. However, basically, the popularity of the cup supported our previous interpretation: People expect direct access to functions through everyday objects; navigation through a menu is considered troublesome.

For most testers, the cup showed arbitrary behavior.

Nonetheless, using the cup led to surprises for most users. Although it was not required for turning the TV on or off, it was nonetheless tried out by all testers throughout other tasks. In total, only nine testers could tell that turning the cup handle was responsible for power management and only four users could actually tell how the cup handle had to be turned for powering the TV on or off (cf. 5.5.2—“Users’ Interpretations”). The video observations revealed that some people had difficulties with the cup and thought it worked arbitrarily. For example, when the TV was already on and the cup was activated and put down with the handle at twelve ‘o clock, the cup actually switched the TV off, which led to surprise. User H.C. said: *“Oh, is it off? But I didn’t do anything!”*. Another tester put the cup down and moved it over the table surface, holding it by its handle. When she discovered that it had no effect, she put the cup away, unintentionally turning the handle to three ‘o clock. Consequently, the TV was turned on right after she picked up the cup again. *“Oh, is it [the TV] on because I put it [the cup] away? That’s strange.”*

Redundancy could lead to state inconsistency.

In the WOZ study, the pattern of turning an object for resembling a classical on/off switch was not observed. We still included this mapping in our interactive prototype for testing redundancy. Tester, H. C., who discovered that both the tissues and the cup were responsible for power management was confused and remarked: *“The cup and the tissues had the same function [i.e., turning the TV on and off].*

This is confusing. A particular function should only be assigned to one object, not multiple.". This was indeed a problem. Firstly, he expected one object to be responsible for a function. Secondly, he encountered state inconsistency: When the cup handle was turned to twelve 'o clock, the TV should be off. However, the TV could be turned on again with the tissues while the cup handle stayed at the previous position. This is a conflict as actuation for everyday objects does not exist. Although redundancy loosens dedication to a specific object-gesture-action mapping by giving the user the freedom to choose between different predefined options, we do not think that it is a good solution as state inconsistencies can occur, which lead to user surprise. All in all, we think the cup failed, as most users were not able to interpret how it actually had to be used.

As regards the intuitiveness of the predefined mappings, Figure 5.19 shows the user-rated results for both scenarios at a glance.

5.5.2 Users' Interpretations

After each scenario had been completed, the participants were asked for their interpretations as regards how each object had to be manipulated and which effect it had on the target system. Logically, the accuracy of the right interpretation correlated with the task success. However, in some cases, additional functionality or misinterpretation were attributed as result of interaction flaws.

Interaction flaws led to misinterpretation.

As regards the television scenario, the function of the wallet was identified by no-one. Pushing anywhere on its surface, the wallet served as universal cancel function for fading out the channel list, the menu, or the volume indicator on demand, i.e., before the automatic fade-out of ten seconds interfered. The testers also disagreed that this mapping was intuitive (Q3f: $M = 1.50$, $SD = 0.82$). Actually, this was not a distinct task and nobody missed this function; automatic fade-out was satisfactory. The reason why people did not detect the wallet function, even though it was pushed frequently, was because the

The function of the wallet was not identified.

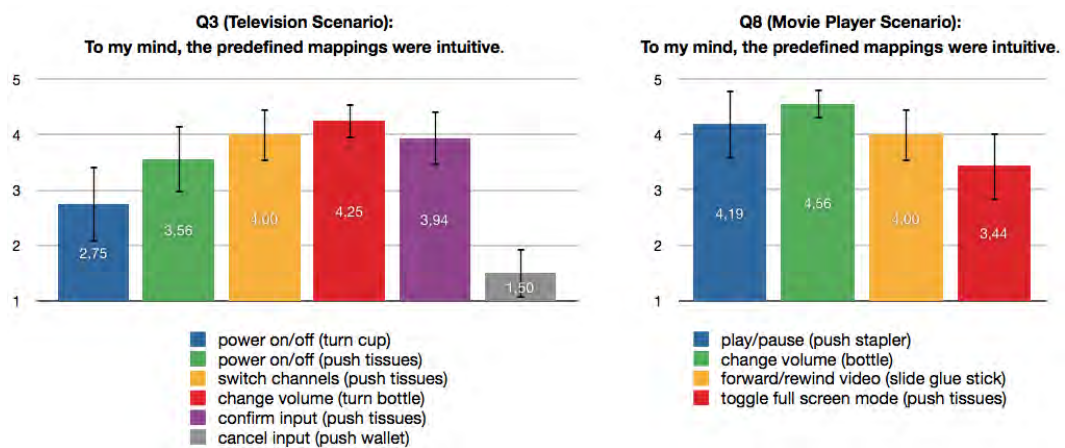


Figure 5.19: User-rated intuitiveness of predefined gestures at a glance. The bars represent the average rating on a five-point Likert scale, 1.0 meaning, that the users strongly disagreed, that the predefined mappings were intuitive, whereas 5.0 means the opposite. Error bars indicate the 95% confidence interval for the mean.

fade-out process could only be noticed when on-screen content was activated by either using the tissues or the bottle. Hence, this required the user to have two objects activated at once; otherwise the layer would have been faded out automatically after some time. Nonetheless, multiple simultaneous object manipulation was practiced by only four participants. Just one of them had the bottle and the wallet activated, but before he pushed the wallet, the volume indicator had already vanished. Basically, this incident showed that people did not consider appropriating objects in combination and therefore matched the findings from the WOZ study.

Originally, the cancel function should have been attributed to the tissues. However, due to limited resolution of the Kinect, the function had to be outsourced. Consequently, it served as universal cancel button, akin the home button of an [iOS](http://www.apple.com/ios)⁹ device. Despite bad mapping, the wallet reinforced the finding that multiple simultaneous object manipulation was rarely practiced.

Multiple simultaneous object manipulation was rarely practiced.

As regards the bottle, all participants could tell that turning it or its cap resulted in volume change, but only four testers stated and experienced that turning the bottle

⁹www.apple.com/ios

– having the right mode activated – also influenced the level of brightness. Yet, among the remaining twelve, three people imagined that the bottle could have played a role for changing the brightness, akin to the volume task. For example, C. L. remarked: *“I imagined the bottle rotation could adjust the brightness like the volume, but it did not.”* (In fact it did, but she did not find the menu.). This showed that the mapping was a logical consequence to the mapping for the volume task.

Adjusting the brightness by turning the bottle was considered a logical consequence of the gesture for changing the volume.

As mentioned before, most people found out, that turning the cup influenced the standby mode of the television set. Yet, only four people were able to tell how the handle of the cup had to be positioned for the TV to respond accordingly. All others attributed arbitrary behavior to the cup – the positioning did not make any sense to them.

Most testers could not tell how to turn the cup for switching the TV on/off.

In the context of switching the TV on, two testers who used the tissues for this task attributed additional functionality to this item. They pushed the upper right corner of the pack of pocket tissues and interpreted that this area was mapped to a virtual power-on switch. Actually, the entire upper part of the tissues was mapped to the up-“button”, which was also responsible for turning the TV on when it was in standby mode. Hence, the two persons attributed an additional button to the tissues, although this was not the case. Both were also pushing the same region again when they were asked to turn the TV off. Of course, the goal was not achieved; the upper right corner did not feature this special power button. Consequently, they were surprised that the TV could not be turned off that way.

By mistake, two testers attributed a power button to the tissues.

Concerning the glue stick, five participants added a secondary function to it. Turning resulted in slight lateral movement (cf. above) which again resulted in flipping of video frames. The testers interpreted this behavior as very fine adjustment of the playback position because the same frames were repeated after a full turn and the QuickTime slider knob moved within a small interval. Actually, this interpretation was wrong. Slider resolution in this task was relatively rough: A total of 640 pixels (depending on the resolution of the web camera) minus a small offset to the left and right yielded a mapping of 600 pixels to a video length of 28:56 minutes with 25 fps. Hence, moving the

By mistake, five testers attributed a second function to the glue stick.

stick by just a pixel to the left or right always resulted in a jump of about three seconds in playback time, known as the slider resolution problem [Hürst and Jarvers, 2005].

The examples showed that interaction flaws influenced the user's interpretation of the predefined mappings. Therefore, it is important to consider filtering out such flaws in order to prevent wrong interpretation.

5.5.3 Mixing Everyday Object Usage with Appropriation

An IUI must be informed when to interpret object manipulation and when not.

Everyday objects are specifically designed tools and are used from time to time throughout the day. The cap of a bottle is turned off for pouring water into a glass when the user needs to drink (cf. Figure 5.20). In addition to this normal object use, appropriation in the sense of IUIs will result in a mix of using the object by design and for utilizing it as a temporary controller. In front of the TV, the user might primarily use the bottle for drinking. Yet from time to time, she might want to use it as volume controller. Basically, a system that allows for appropriation of everyday objects cannot distinguish between normal and appropriated usage of an object. In the bottle scenario, when the user screws off the cap to drink water, the system will recognize the gesture – and interpret it as volume change. Thus, a secondary “goal” is achieved, even though that was not the user's intention. Hence, the user has to tell the system when to stop interpreting object manipulation.

Activation and deactivation gestures were inspired by the WOZ study.

Inspired by the observations and the feedback received from the previous study, we decided to make use of special gestures for *activating* or *deactivating* an object from the system. On the one hand, whenever the object was activated, manipulation was interpreted by the system. On the other hand, an object being in deactivation mode appeared invisible to the system, i.e., any interaction with the artifact was not interpreted. Recalling the example, when a user wants to turn the cap off the bottle for drinking, the object should be deactivated; if she wants to control the volume, she must activate it beforehand.



Figure 5.20: Example of an interrupt. After having accomplished a task, the user was asked to perform a usual task with the previously appropriated item, such as pouring water into a glass after the bottle was used for volume control.

For object activation, we used an explicit pointing gesture: The user had to point the object at the target display and hold it in that position for three seconds. A tone indicated object recognition and activation. Deactivation was achieved by either pointing again (thus, a toggle mode) or by hiding the object. The latter consisted of two options: The user could either put the object outside the interaction space which implicitly deactivated the object after three seconds. The alternative was explicit deregistration which was achieved by covering the object with a hand for three seconds.

Activation was explicit, whereas deactivation was achieved by either using an explicit or an implicit gesture.

The pointing gesture was used because it was observed in the WOZ study and suggested by the testers. Deactivation by pointing was meant to help the user being aware

Explicit deactivation was meant to prevent side effects caused by movement of an object.

that she had deactivated a certain object. As mentioned before, there were no visual cues as to the object state; only audible feedback was given just in time. As regards hiding, we added explicit deactivation by hand to give the user the option of deactivating an object without having to lift or move it. This was particularly important where movement of an object was linked to an action. Otherwise, the deactivation process itself would have been interpreted as movement and caused the system to change the current state which was obviously not intended in this case.

Appropriation of objects was mixed with normal object usage.

Activation by pointing was the same for both scenarios. As regards deactivation, hiding and pointing were counter-balanced: One half of the participants used pointing in the TV scenario and hiding for the QuickTime Player tasks; the other half did vice versa. For realistic testing of activation and deactivation we mixed scenario-related tasks, i.e., system control with normal object usage, such as using the bottle for controlling the volume and for drinking. The sequence of these interrupts and their concrete goals are listed in Table 5.1 for the TV scenario and in Table 5.2 for the movie player, respectively. Side effects that could happen when the object was not deactivated while performing the interrupt are included.

Pointing as activation gesture was considered a natural approach.

Pointing as activation gesture worked promisingly and was consistently applied by the testers whenever they thought the object they wanted to use was not activated (anymore). The testers almost all agreed that pointing was a natural approach (Q13a: $M = 3.75$, $SD = 0.68$). In addition, having to perform an activation gesture before the appropriation of that object was found neither annoying nor comfortable. However, a deactivation gesture was considered slightly more annoying than an activation gesture. (Q12b: $M_{deactivation} = 3.50$, $SD_{deactivation} = 0.97$, Q12a: $M_{activation} = 3.19$, $SD_{activation} = 1.11$).

As regards the two deactivation methods, we wanted to find out whether specific deactivation patterns existed. Moreover, we were interested in whether side effects were a concern depending on the method and whether the users were aware of the actual object state (activated vs. deactivated).

Wrong object state, i.e., the user thinking that the object was activated when in fact it had been deactivated before, was identified 15 times for pointing and twelve times when hiding was used (summed up for all users for all tasks). However, wrong object state through pointing was detected eleven times by just one person. The toggle mode confused her. First, she registered an object and tried it out. Then she put it away and activated another object. Later, when she decided to use a previously used item again, she pointed first, because she failed to remember that the object was still activated. The second pointing, however, caused object deactivation – any gesture was ignored, which surprised her. Implicit deactivation by putting the object away (hiding metaphor) would have solved this problem. In fact, for the hiding method, which followed later, she experienced no problems at all. Identifying this problem as outlier, we can say that pointing was less irritating for the tester for identifying the object state. This correlates with testers' feedback: They agreed to being more aware of the object state when using pointing (Q16a: $M = 3.75$, $SD = 0.93$) than hiding (Q16b: $M = 2.81$, $SD = 1.28$) for deactivation. Our explanation is that pointing is explicit whereas hiding is implicit. The user had to hold the object and point it at the system – it made her aware of the deactivation process, which was not necessarily the case when simply putting the object aside. Hiding was used implicitly by eight users; the remaining half applied explicit hiding, i.e., by covering the object by hand.

Pointing made the users more aware of whether an object was activated or not compared to hiding.

Despite better performance for pointing, hiding was preferred by most testers (Q15, cf. Figure 5.21). The reason for this could be that hiding felt slightly more natural to the users (Q14: $M = 3.50$, $SD = 1.10$) than pointing as deactivation gesture (Q13b: $M = 3.31$, $SD = 1.01$). Besides, the option of implicit deactivation required less effort and was therefore perceived more comfortable.

Despite better performance for pointing, users preferred the hiding metaphor.

However, wrong object state was not only due to the user's memory. It was majorly influenced by *unintended deactivation*. Unintended deactivation could happen in two cases: (1) when the user used an object and unintentionally hid the white marker with the palm, as this was interpreted

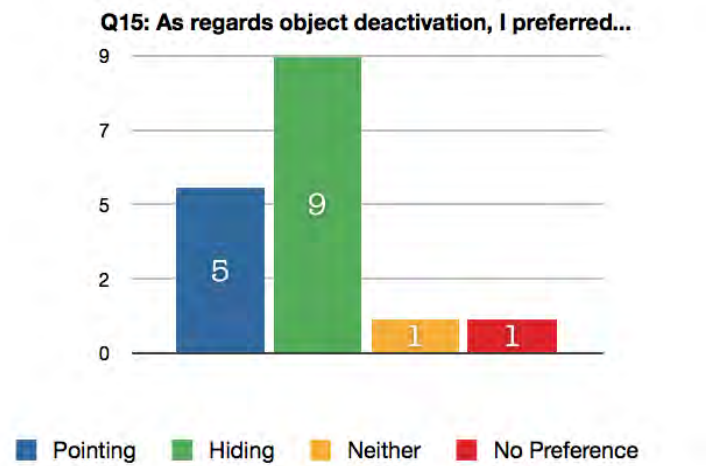


Figure 5.21: Histogram for deactivation preference. The hiding metaphor was preferred by most users as is allowed for implicit object deactivation. Nevertheless, five out of 16 users preferred explicit pointing which made them more aware of which object they had deactivated.

The hiding metaphor contributed to unintended object deactivation while the user was interacting with the artifact.

as explicit hiding gesture and (2) when an activated object was covered by another one that the user was currently activating. In this case, the first object was deactivated as explicit hiding was suggested again. In total, unintended deactivation happened 37 times in the TV scenario, but only eight times for the movie player tasks. Unintended deactivation could not happen for the pointing method as it is explicit.

Unintended deactivation led to surprise and frustration.

Unintended deactivation happened much more frequently in the television scenario due to the pocket tissues. When the users pushed the up-“button”, they automatically hid the white marker in the middle of the object surface. The system interpreted this as hiding gesture. Consequently, the users were frustrated and wondered why the tissues reacted to pushing first but then – all of a sudden – ignored all input. Finally, they gave up. Therefore, the hiding gesture was inappropriate for gestures that result

in covering most of the object surface. Similarly, the bottle was responsible for unintended deactivation. As it is large, activating it caused previously activated objects to be deactivated: The bottle hid the white marker of some objects put in the interaction space before. Yet, multiple object use was not frequently observed and therefore, case (2) did occur less frequently than case (1).

Yet, the examples showed that explicit hiding was no good solution, at least when the major part of the object surface was covered by the palm while the user was interacting with the object: The system misinterpreted the interaction. Yet, explicit hiding was beneficial as it prevents side effects when switching to normal object usage.

All in all, explicit hiding caused too many interaction problems.

Side effects occurred 16 times when the pointing method was used and 13 times when hiding was used for deactivation. As these occurrences were nearly equal in general, for the scenarios they were not: Side effects were observed nine times in the TV scenario and 20 times in the movie player scenario. This notable difference stemmed from the involvement of a horizontal movement gesture in the second scenario. When the users were asked to put glue on a piece of paper after having forwarded the video with the glue stick, people were unaware that grabbing the object would cause the video position to change, as the object had not been unregistered until then. When the pointing gesture was used, a side effect was inevitable. By applying the hiding gesture, however, the user could just hold her hand over the stick and wait until the deactivation tone was audible. Grabbing the object afterwards had no effect on the playback position. Yet, this “trick” was rarely used. A similar problem could be observed for the bottle. When grabbing the item resulted in a slight rotation of the object itself, the volume changed whenever the object had not been deactivated by holding the hand over it for a while. Yet, when the tester quickly moved the bottle vertically, accidental turning was too small to be recognized. Consequentially, no side effects occurred.

For objects that were linked to gestures of movement, implicit hiding caused side effects.

All in all, the hiding gesture resulted in fewer side effects. Yet, eight participants predominantly put the objects out of the interaction space instead of covering them. Hence, side effects still occurred.

As the stapler was pushed with different intensity when it was appropriated compared to usual stapling, the system was able to anticipate the user's intention without the need for a special activation or deactivation gesture.

Interestingly, as regards the stapler, no special deactivation gesture was necessary. Most people used the stapler differently for the two kinds of usage: When it was appropriated as a button, the testers pushed the device softly with the forefinger. For stapling paper, however, they applied much more pressure by putting the fist on the stapler. Participant, H. R. remarked: *"The stapler [he meant the system] did not react when I stapled the paper, but only when I wanted to play the video."* For both ways of manipulation, the system behaved as designed: It detected different depth values. The fist appeared closer to the camera than the tip of the forefinger on the stapler. Of course, this effect also made activation superfluous. Yet, for reasons of consistency, the stapler had to be activated by pointing. Similarly, electronic devices could be uncoupled from special de-/activation gestures. As long as the electronic device is turned off, it can be appropriated; when it is turned on, normal use is desired.

Deactivation Patterns

Counter-balancing pointing with hiding as deactivation methods for two scenarios resulted in four different user constellations, consisting of eight testers each. We analyzed the user's behavior regarding when she deactivated objects. A summary of the patterns observed including their distribution among the users is given in Table 5.4.

After a task was accomplished, the users "forgot" about object deactivation.

Interestingly, each activation method showed a nearly consistent distribution of the pattern types regardless of the scenario. As regards pointing, most testers almost always deactivated an object directly after they did not need it any more. Yet, when the task (and therefore the goal) was accomplished, they forgot to deactivate it. This seemed to happen because the prior intention, i.e., the goal, had already been reached. Likewise, Norman [2002] argues that a user forgets about circumstantial tasks when the goal has been accomplished. Nevertheless, when being asked to accomplish an interrupt task, the testers deactivated

Pattern	Scenario	
Deactivation by Pointing <i>was done...</i>	Television	Movie Player
	never	2
	immediately	5
	arbitrary	1
Deactivation by Hiding <i>was done...</i>	Television	Movie Player
	always	4
	arbitrary (object put aside)	1
	always	3
	arbitrary (object covered by hand)	–

Table 5.4: Deactivation patterns observed by quantity. For both deactivation methods the distribution was nearly equal for the two scenarios (± 1 user, $N = 16$). As regards pointing, most testers immediately deactivated an object before using another one. Patterns for hiding showed no preference.

the object first, as they were aware that side effects could occur. As the interrupts were posed directly after a task had been solved, the users were thus still aware of the recent object being still activated. We believe, however, if the interrupts had been performed after a break (e.g., after watching television for some minutes), the users would not have been as conscious about the object state.

A second pattern that could be observed was that people did not care about deactivation at all. When the artifacts were not needed anymore, they put them aside often leading to wrong object state and side effects. Arbitrary deactivation was considered only twice for the TV scenario and only once for the QuickTime Player scenario. The users who followed this “pattern” sometimes deactivated an object after usage and sometimes they did not.

One group of testers did not care about deactivation at all.

As regards the hiding gesture, immediate deactivation of an object before picking another one was consistently observed for both scenarios. In fact, it was not possible to forget to deactivate an object when it was put aside the workspace which was done by four people per scenario. Explicit deactivation by hand was the second pattern observed and practiced by four/three persons per scenario. For this method, both implicit and explicit

Covering an object by hand for deactivating it could lead to misinterpretation by the system during interaction.

deactivation could lead to interaction problems. Whereas covering the object by hand could be used to prevent side effects that occurred when the object was put away, this technique was problematic as it was responsible for unintended deactivation when pushing the tissues or activating the bottle (cf. above). In fact, we could also observe that users who were attributed to hiding in the TV scenario found channel browsing with the tissues less easy (Q1a: $M_{hiding} = 2.88$, $SD_{hiding} = 0.83$) compared to users who were assigned to use pointing for this task (Q1a: $M_{pointing} = 3.63$, $SD_{pointing} = 1.19$). On average, however, the rated ease of handling the tasks for each scenario was nearly equal (Q1, television tasks: $M_{pointing} = 3.41$, $M_{hiding} = 3.50$, Q6, movie player tasks: $M_{pointing} = 4.10$, $M_{hiding} = 4.03$).

Two users
experienced object
competition.

As a supplement to these patterns, only five testers put two or more objects in the workspace at the same time. Yet, this behavior was observed only once or twice per person. For three users, multiple activated objects led to conflicts. The first conflict was *state inconsistency* with the cup and the tissues that could both be used for switching the TV on/off, as described in section 5.5.1—“Task Performance for the Television Scenario”. The second problem was that of *object competition* which was observed twice for the glue stick in the QuickTime Player scenario. Both users did not deactivate the glue stick when they added the stapler to the workspace. When they pushed the stapler only very short playback could be noticed. The users were surprised as they expected the video to be played back normally from the latest time position. The problem was, that one object concurred with another: As the system reacted continuously to movement of the glue stick, it bound the video playback time to the corresponding position of the glue stick. When the stapler was pushed, playback was started, but the glue stick then told the system to go back to the initial time position. The solution was to set the glue stick to the desired time position, deactivate it, and then use the stapler for playback; this way the glue stick could not interfere anymore. This solution was found by both testers. Yet, it took them some time to understand why playback was not possible at first.

For such constellations, it is therefore important to define object policies the system must follow. For the situation explained above, the stapler should be assigned a higher priority to prevent influence by the glue stick. Yet, this leads to state inconsistency, as the glue stick is still standing at the initial position, but playback time is changing.

Policies could help preventing object competition despite state inconsistency.

5.5.4 Overall System Evaluation

Next to task-related feedback discussed earlier, we also asked the testers for evaluating the prototype as a whole.

First of all, we were interested in whether the size of the interaction space was appropriate for object manipulation. The size depends on two factors, i.e., on the resolution of the Kinect and the web camera (each 640x480 pixels²) and their distance to the table surface. For both scenarios, the size of about 58cm x 44cm (width x height) was considered being in the range of ideal and too small (Q2: $M_{TV} = 2.56$, $SD_{TV} = 0.51$, Q7: $M_{MoviePlayer} = 2.94$, $SD_{MoviePlayer} = 0.44$, whereas 5.0 means “far too big”). Interestingly, the size was rather considered being too small for television control compared to movie player navigation, although only the latter demanded planar horizontal movement; we thought that a bigger size of the interaction space would be in demand when objects needed to be moved. A possible explanation why the result is slightly different from what we expected could be the higher difficulty perceived for solving the TV scenario. In fact, people tried out objects more intensively compared to the movie player scenario which was considered being more intuitive as regards predefined mappings. There were five participants who put multiple objects in the workspace at the same time. Four of them contributed to the lowest rating for the interaction size. As these results were not satisfying, cameras with higher resolution should be used in future prototypes (cf. 6.2—“Future Work”).

The size of the interaction space was considered too small.

As regards the interaction comfort of everyday objects used as controllers in comparison to dedicated input devices, the comfort felt slightly higher for the movie player

As regards
interaction comfort,
everyday objects do
not compete with
dedicated tools.

scenario (Q9: $M = 2.5$, $SD = 0.73$) compared to the TV scenario (Q4: $M = 2.31$, $SD = 0.72$) which correlated with perceived ease and intuition of gestures. Nevertheless, for both scenarios, interaction was obviously perceived less comfortable compared to dedicated input devices. This is not surprising as dedicated input devices are *designed* for the task environment, but everyday objects are only meant to serve as temporary substitutes in case the original device is unavailable. Yet, for both scenarios the testers agreed (Q5: $M_{TV} = 4.06$, $SD_{TV} = 0.77$, Q10: $M_{MoviePlayer} = 4.19$, $SD_{MoviePlayer} = 0.40$) that they could imagine using the everyday object controllers in case remote control, mouse, and keyboard are missing. This result is promising and motivates for continuing IUI research.

Users needed more
eye contact with the
artifacts than
expected.

When the participants were asked whether they needed eye contact with the object while manipulating it, they neither agreed nor disagreed (Q11: $M = 3.13$, $SD = 1.20$). This result was not satisfying since the idea behind Co-Optjects is to exploit physical affordances of everyday objects for utilizing them as *haptic* controllers. Mapping multiple Multiple virtual buttons to a continuous flat surface, like the tissues-based remote remote control, should therefore be avoided. However, since we were not able to implement all patterns as observed in the WOZ study due to technical limitations, we are confident that using more sophisticated technology will help exploiting physical object properties better and therefore contribute to better eyes-free interaction.

As regards object activation and deactivation, the time span of three seconds was considered too long. Therefore, that time span should be reduced but still be long enough to prevent accidental activation, e.g., when lifting up a cup for drinking.

Testers demanded
for visual feedback
indicating whether an
object was in
appropriation mode
or not.

Using visual feedback for indicating that an object was activated was highly demanded by the testers. On average, they agreed that they would have needed visual feedback for telling the current object state of objects that were currently not occupied by the user's hands (Q17: $M = 4.00$, $SD = 0.97$). These results can be confirmed by the findings of our video analysis. The testers often asked the principal investigator whether a certain object that they

had not used for a certain time was still active or whether it had been deactivated by them before. Hence, purely audible feedback (as used in our study) was helpful during the process of activation and deactivation but not sufficient for the users for being aware of the object state at any time.

5.6 Lessons Learned

Having tested everyday object interaction with an interactive prototype, we could confirm that predefined object-gesture-action mappings based on the patterns observed in the WOZ study worked well. The glue stick which was meant to be slid horizontally for changing the playback position, was an exception. Although the testers agreed that the mapping was intuitive, they were influenced by a previously used object that roughly resembled the stick shape. Hence, having experienced a working gesture for an object of a specific shape, it is likely that the user will execute this gesture again on similarly shaped items.

Predefined mappings which were conform to the WOZ study, worked well.

We tested everyday objects as controllers for two scenarios. The tasks from the television scenario were considered less easy to handle compared to the movie player sample. This finding correlated with the predefined mappings: The testers considered the mappings for the movie player more intuitive.

Users considered the mappings for the movie player more intuitive.

For imitating a realistic workaround, we mixed appropriated object usage with normal object use by posing so-called interrupts to the user. In order to avoid side effects, i.e., for preventing the system to interpret object manipulation during interrupt tasks, special gestures were introduced for activating or deactivating an object from appropriation mode. We tested two different versions for deactivation, called pointing and hiding, whereas the first one was always used for activating an object.

We tested two different methods for object deactivation.

Hiding was preferred over pointing as regards deactivation, since it was implicit. If the user would forget

Hiding as deactivation gesture was preferred over pointing, as it was implicit.

to deactivate an object, it would be decoupled from the system automatically, the object being put aside. In fact, users tended to put an object aside when they were done with it, as most of them directly deactivated the object before grabbing another item; multiple simultaneous object manipulation was rarely observed.

We identified interaction flaws.

With the aid of video analysis, think aloud recordings, and feedback forms, we learned which interaction problems occur in the field of IUIs:

- Using audible hints only, the user can have forgotten which state (activated vs. deactivated) an object is in. Pointing as toggle mode for activation and deactivation then leads to user's surprise when she thinks the object was deactivated, but in fact remained activated. This gesture then deactivates the object and ignores any object manipulation, although the user expects the opposite reaction.
- Although covering the object by hand helps to reduce side effects during deactivation when motion-based gestures are linked to the corresponding object, it can also lead to unintended deactivation. This can happen when the user's palm covers too much of the object surface while manipulating it. Likewise, large objects can hide smaller objects during the activation process. This again surprises the user as sudden deactivation ignores any gesture applied.
- Object competition is also a concern in cases where multiple objects are put in the workspace and are all activated. One object can influence the behavior of another which is not expected by the user and surprises her.
- A function should always be directly linked to an object. Activation of modes through menus is not expected.

Basically, we see potential in IUIs. The study conducted showed that the concept basically works despite interaction flaws that have to be eliminated in the future. These flaws led to users' dissatisfaction, such that they

rated everyday object interaction as less comfortable than manipulation by dedicated devices. This result is not bad at all since (1) these objects are not meant to be used as permanent replacements and – more promising – (2) the users agreed that they could imagine using everyday objects as substitutes for controllers in case dedicated tools are missing.

Users could imagine IUIs as workaround in cases where dedicated input devices are not available.

The idea of using predefined gestures worked but was not as good as expected. Using the right object with the right gesture right at the beginning of a task was observed, but in most cases, the users needed several tries until a task was solved, if not aborted. The question that remains thus is whether predefined gestures actually contribute to more *instant* and *spontaneous* access to IUIs comparing to an end user programming approach. Therefore, in a future quantitative study, it will be interesting to investigate whether trial and error takes less time than individual programming of an object before being able to use it (cf. 6.2—“Future Work”).

In most cases, users needed several tries until the right mapping was found.

Alternatively, redundancy could be used for supporting various mental models. This approach gives the user more freedom in appropriating objects; she can pick the object that fits best to her imagination. Thus, providing multiple options for predefined object-gesture-action mappings enhances the probability, that it matches the user’s mental model so that the time needed for trying out several objects (before the right one is found) is reduced. Therefore, redundancy can be considered a solution to the second error source, as listed in section 5.5.1—“Users’ Performance on Tasks”. Nevertheless this workaround contributes to state inconsistency when multiple objects are attributed the same function, as experienced with the cup and the tissues.

Redundancy could help reducing trial and error.

Based on the results of this study, we propose to combine explicit pointing with implicit deactivation as solution to the third error source (i.e., the interplay of normal and appropriated object usage). As implicit hiding does not cancel out side effects occurring due to motion-based gestures, we need to investigate this problem in future prototypes. In order to keep the user informed about which objects are activated, we propose a visual indication

As regards voluntary and non-voluntary object usage, we propose a combination of explicit object activation and implicit deactivation.

per object, e.g., by projecting a small icon onto the artifact surface. Audible feedback should be kept though, because it enables the user to immediately start blind interaction right after deactivation; visual indication is thus provided for recalling the object state at a later point in time.

We see the
foundation for IUIs
being laid.

Finally, we see the first steps made towards IUI research. Using everyday objects as haptic controllers principally works, but interaction flaws remain. The results of this study are not completely satisfying but motivating. Therefore, we present our plans for future research (cf. 6.2—“Future Work”) including a list of recommended guidelines (cf. 7—“Instant User Interface Guidelines”) based on the results of all studies presented in this thesis.

Chapter 6

Summary

6.1 Summary and Contributions

This work investigated the appropriation of everyday objects as haptic input controllers for interactive systems. The topic has been inspired by appropriation applied on objects used in everyday life, such as a knife which can temporarily substitute a letter opener. Sharing the same physical affordances, a quotidian object (the knife) helps becoming less dependent on a specific tool (the letter opener). We transformed this methodology from passive world to interactive systems by introducing a new interaction metaphor, called *Instant User Interfaces* (IUIs).

We introduced a new interaction metaphor, called *Instant User Interfaces*.

Before investigating spontaneous appropriation of artifacts, we inquired the ubiquity of objects: Which objects are actually within reach in everyday life? We conducted a diary study with 19 participants and asked them to take pictures of objects within their reach from time to time on a work and on a leisure day. The photos were taken in offices, kitchens, bathrooms, living rooms, bed rooms, or outdoors. By counting and categorizing objects visible on the pictures, we achieved a top ranking representing the most frequent objects at hand while being in the office or while spending free time. The top five consists of containers/boxes, paper, bottles, drinking vessels, and pens.

We conducted a diary study for inquiring the ubiquity of objects.

We conducted a Wizard of Oz study for investigating how users appropriate objects.

Having identified *which* objects are ubiquitous, we used this ranking to investigate how people would actually appropriate quotidian artifacts for using them as controllers by exploiting physical affordances. Therefore, we tested everyday object interaction for interactive systems in a Wizard of Oz (WOZ) study, in which testers were asked to perform gestures of their choice on objects they found most suitable for solving simple tasks assigned in various scenarios. Activities were related to watching television, playing a video game, holding a presentation, or controlling simple desktop GUIs, to mention a few. We did a qualitative analysis on object-gesture-task mappings suggested by the users and identified basic patterns on object appropriation for the scenarios given.

One basic methodology of appropriation observed was *Physical Instantiation*.

Basically, people were looking for button-equipped objects for replacing original remote controls, therefore trying to imitate the original tool as well as possible. For changing the focus from one system to another, the users either pointed an object at the target or they used a separate object, e.g., for changing the focus by turning a cylindric item. As regards continuous input, people tended to instantiate the target by using an object that represents it. Manipulation on the object was meant to have the same effect on the remote target system. We denoted this paradigm by *Physical Instantiation* (PI), in which the user does not try to imitate a separate remote control for the system but rather uses a physical instance that represents the *entire* system. Particularly for controlling a slider or a rotary button, PI was most frequently applied. Furthermore, people considered the shape of an object important and exploited physical idiosyncrasies such as the handle of a cup representing a clock hand for control. In the cases where PI was not applied, one to two objects were mostly sufficient for controlling basic tasks of a remote system.

We developed an interactive prototype being capable of turning passive objects into input controllers.

Although the testers were not limited in their choice of gestures, the WOZ prototype only supported incremental feedback, i.e., the new system state was visualized only *after* a task had been completed. Consequently, the distinction between voluntary and non-voluntary interaction with the system through object manipulation could not be tested. Therefore, we developed and evaluated

an interactive prototype, technically based on a combination of a depth camera and optical tracking of fiducials attached to objects. Using this technology, passive objects were turned into active controllers capable of detecting pressure-sensitive touch/deformation, 2D movement, and circular rotation. In a qualitative user study, we tested predefined object-gesture mappings adopted from the WOZ study combined with two methods for indicating voluntary and non-voluntary object usage to the system.

Our contributions are the introduction of a new interaction metaphor, called IUI, alongside with qualitative investigation on everyday objects and their appropriation for controlling sample systems. Generally, we see potential in the field of IUIs: The idea is appreciated by the testers and considered a useful solution in case a special input device is missing. Yet, only simple tasks, such as watching TV or navigating a non-complex GUI with a low number of UI widgets, seem to work for this metaphor. General interaction patterns for object appropriation in the sense of IUIs exist despite the huge ways of how to appropriate everyday items for using them as controllers. However, sophisticated technology such as reliable automatic object detection and tracking combined with highly precise gesture recognition is needed for fluent unlimited object interaction.

We introduced a new interaction metaphor alongside with qualitative analysis of IUIs.

We see the foundation for IUIs being laid. The next step will have to be to concentrate on sophisticated technology before conducting further profound qualitative and quantitative studies (cf. 6.2—“Future Work”). Future implementation and studies should be based on the guidelines that we propose to follow based on the results of our studies (cf. 7—“Instant User Interface Guidelines”).

We see the foundation for IUIs being laid.

6.2 Future Work

Before being able to conduct more studies in the field of IUIs, we need to investigate sophisticated technology. Strictly speaking, it needs to fulfill the following requirements:

Sophisticated technology is needed.

- As a basis, we need automatic and reliable object recognition. Ideally, it is able to create a 3D model from the artifact shape that is used for identifying the object. Automatic object identification will foster spontaneity as it eliminates the need for preparation of everyday objects with fiducials before being able to appropriate them.
- In addition, we require real-time tracking of everyday objects so that the user can freely interact with the artifact while holding it in her hands, as known from interaction with dedicated tools. As the user moves her hand or entire body, so does the artifact, making it necessary to inform the system about the position change.
- Furthermore, we need precise and sensitive gesture recognition, so that manipulation of small object parts is supported. This includes the recognition of pushing small-sized buttons, for instance.
- Finally, three-dimensional input recognition is mandatory. Apart from the top surface of an object, manipulation of lateral surfaces of the artifact must be recognized, too.

Sophisticated technology will be necessary for conducting further studies.

If these technical requirements are met, we expect unrestricted and consequently more natural and satisfactory interaction for the user. Unlike our interactive prototype, there would be no need to put an object on the table and re-calibrate the system after an object has been moved. Furthermore, even small items, such as buttons of a pen or mobile phone could be linked to gestures, as observed in the WOZ study. Hence, sophisticated technology enables the conduction of further studies meeting more natural circumstances. Therefore, we are planning to re-conduct the latest study.

We will to conduct studies comparing the performance of input by dedicated tools vs. input by appropriated objects.

Besides study re-design, we are planning to conduct comparative studies. The first one will contrast interaction with everyday objects with control via dedicated input devices. This enables us to analyze input precision of everyday objects and to what extent they keep up with dedicated input devices. Moreover, it will help us to

understand what effects everyday objects will have on the role of dedicated input devices. Do constellations exist, in which the user makes no difference between using a dedicated tool or an appropriated object? The results of this study will strongly depend on the feasibility of the technology used.

The second (quantitative) study is meant to investigate spontaneity. Can a system based on predefined gestures provide quicker access to IUIs than a system that is based on end user programming? Regarding the first approach, the user probably needs some time until the right preset object-gesture-action mapping is found, whereas the second system will cost the user time until an individual mapping is programmed. Which approach is faster? Can redundancy, i.e., the provision of multiple predefined gesture mappings for the same action, help to reduce trial and error for finding a working mapping that matches an individual's mental model? To what extent will redundancy influence user experience negatively, e.g., by introducing state inconsistency between different object due to missing actuation?

We are planning to conduct a qualitative study as regards aspects of spontaneity supported by IUIs.

Furthermore, we will examine our previously posed hypothesis. We claimed that – for a set of common controls – people will link the object with the strongest affordance (e.g., the object that is most likely to be pushed) to the main function of the interface.

Finally, we have to investigate canceling out various interaction flaws, such as object competition, which can result in taking away control from the user as one object influences another and vice versa. We also need to examine implicit deactivation methods for uncoupling an object from an IUI without causing side effects during the deactivation process.

We will investigate canceling out interaction flaws.

For those who want to investigate the field of IUIs (thus also for aspects listed in future work) we recommend considering the guidelines towards IUI research, listed in chapter 7—“Instant User Interface Guidelines”. It presents basic recommendations based on the outcome of your qualitative studies for supporting IUI design.

We present a list of guidelines for IUI research.

Chapter 7

Instant User Interface Guidelines

This chapter presents a list of 17 recommendations to follow when dealing with research in the field of IUIs. The advice is based on observations made in two qualitative user studies (cf. 4—“Wizard of Oz Study: Interaction Patterns for Instant User Interfaces” and 5—“Testing Everyday Object Interaction with an Interactive Prototype”) towards appropriation of everyday objects as temporary controllers.

7.1 Recommendations and Guidelines

1. *Consider appropriation a temporary mode.*

The most frequent mode of use for an everyday object is the use it is designed for. A pen button, for instance, will be used most often for pulling the pen refill in and out for writing and not for imitating a toggle button. Therefore, the designer must always keep in mind that appropriation is of only marginal use.

2. *Appropriation coexists with dedicated input devices.*

A dedicated input device is specially designed for being able to handle a set of tasks as well as possible. Everyday items used as controllers should always be considered a compromise, only temporarily substituting the original control (→ Rule 1). As these items are not specially designed for the task at hand, they do not compete with the original tool.

3. *Exclude objects of soft texture.*

People consider soft materials such as textiles inadequate for manipulation. A certain rigidity is expected though deformable objects are fine. People prefer lightweight and small items for appropriation.

4. *Exploit physical affordances.*

Always map virtual controls to clearly physically perceptible and distinguishable parts of an object. Avoid blind mapping. A smooth surface represents one control and should not be logically subdivided.

5. *Support recognition of pressure-sensitive touch, squeeze gestures, and three-dimensional stationary and non-stationary movement.*

People consider these gestures a basis for IUI interaction. Stationary movement refers to motion-involved manipulation on the object surface.

6. *Use consistent gestures for objects looking similar.*

For different objects that look physically similar, e.g., by sharing the same shape, always map the same types of gesture. Once the user has explored a certain gesture that can be applied to the object, she will expect an object looking similar to react to the same gesture.

7. *Be aware of gestures executed involuntarily.*

Due to nervousness or out of habit, a user might play around with objects without intention. Objects, such as a pen, are moved, grasped, or pushed thoughtlessly without the user being aware of this. Therefore, unintended application of gestures must be filtered out, e.g., by requiring the user to actively point at the target while appropriating the object (→ Rule 13).

8. *Avoid gestural side effects.*

An executed gesture might be accompanied by inevitable execution of other gestures, like turning is followed by slight lateral movement. Gestures executed accidentally must be filtered out in case they are linked to a different action.

9. *Link coherent functions to one object.*

Multiple, simultaneous object use is rare. It is not likely that people will use multiple objects in combination. Therefore, map coherent functions to the same object, without overloading it.

10. *Provide direct access to a function.*

Looking for modes or deep navigation through menus using everyday objects is considered tedious. Therefore, link a function directly to an object so that it is executable by a single gesture.

11. *Mind inconsistent object states.*

Due to missing actuation, state inconsistencies can occur when one object influences the state of another object when redundancy is applied. Keep in mind that actuation is no option to synchronize object states, since everyday objects should not be altered physically.

12. *Provide policies for concurrent objects.*

In order to avoid that one object can influence the state of another object and vice versa, stick to → Rule 9. If not possible, provide policies such that the user's latest action is properly executed and not influenced by another object. However, keep → Rule 11 in mind.

13. *Continuous input should conform to Physical Instantiation.*

In a case where continuous input is to be controlled with everyday objects, it is likely that people will apply Physical Instantiation. This is also valid for complex interfaces.

14. *Use explicit activation for entering mode of appropriation.*

To make people currently aware that the object will be in appropriation mode, let them execute an explicit gesture. Pointing is a good choice.

15. *Use implicit deactivation for exiting mode of appropriation.*

As explicit deactivation dissatisfies the user, use an implicit method, such as automatic deactivation when the object is put aside. If possible, exploit interaction differences between normal object usage and appropriated use for eliminating the need for a special activation and deactivation gesture.

16. *Use visual feedback for indicating mode of appropriation.*

At any time, the user must be able to check whether an object is in appropriation mode or not, as she might have forgotten the state after some time. Use visual indication *on* each object, e.g., by projecting an icon on the object surface indicating the state.

17. *Do not map critical tasks to everyday objects.*

As everyday objects feature less precise input compared to dedicated tools, and accidental execution of gestures linked to irrevocable actions can happen, critical tasks should not be linked to appropriated objects.

Appendix A

Appendix for the Wizard of Oz Study

Appendix A contains additional material used for the Wizard of Oz study. It includes all sketches of the scenarios presented, as well as the feedback form handed out to participants.

A.1 Scenario Sketches

The sketches of the scenarios presented in the study including all visual updates are listed subsequently. Tasks were faded in one after another. Only *after* a task was performed, the slide was updated. The images are arranged in columns (left to right), each to be “read” from top to bottom.

A.2 Feedback Form

In total, the feedback form included 16 questions.



Figure A.1: A1: Holding a Presentation

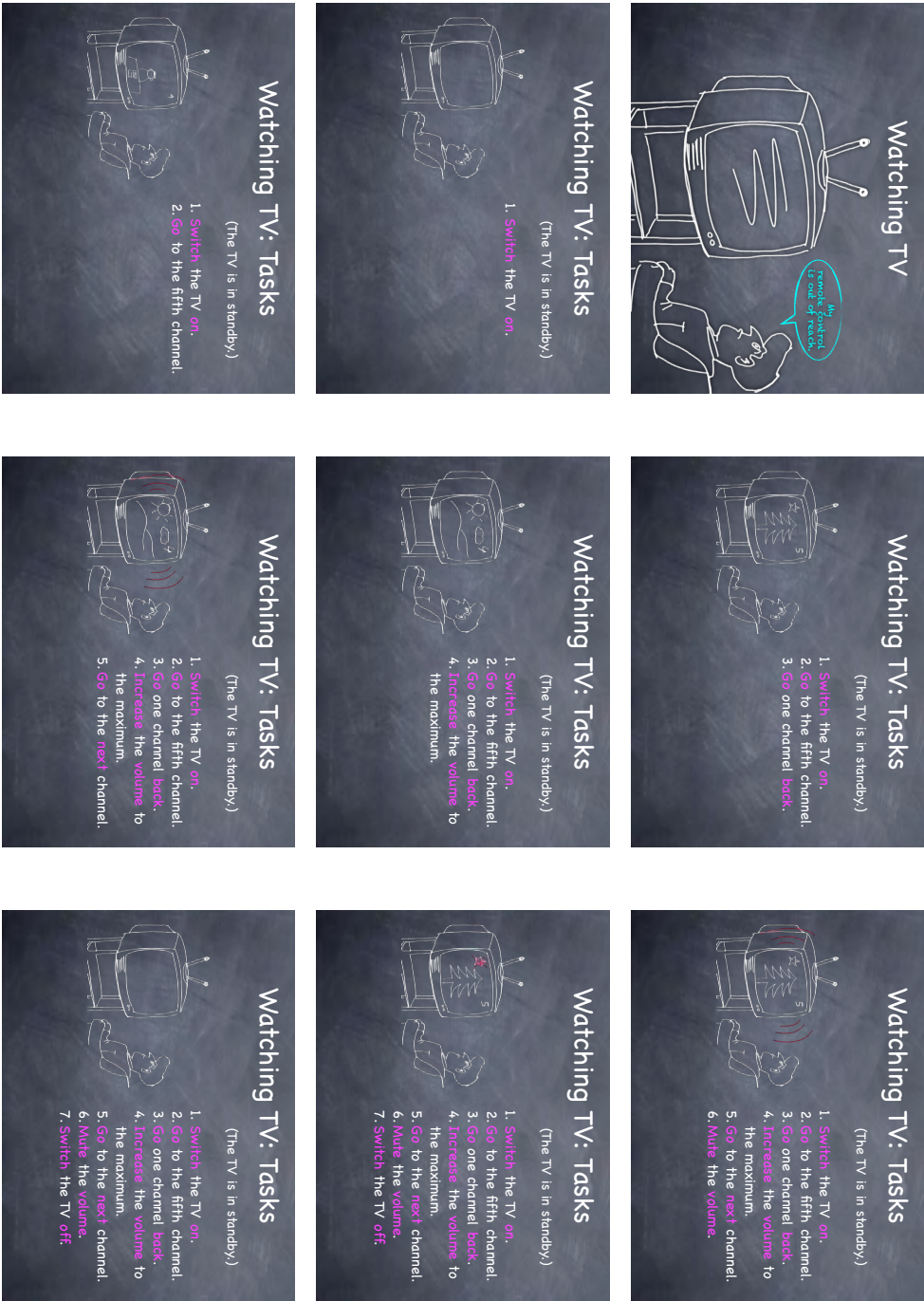


Figure A.2: A2: Watching Television



Figure A.3: A3: *Playing a Video Game*



Figure A.4: A4: Operating the Light

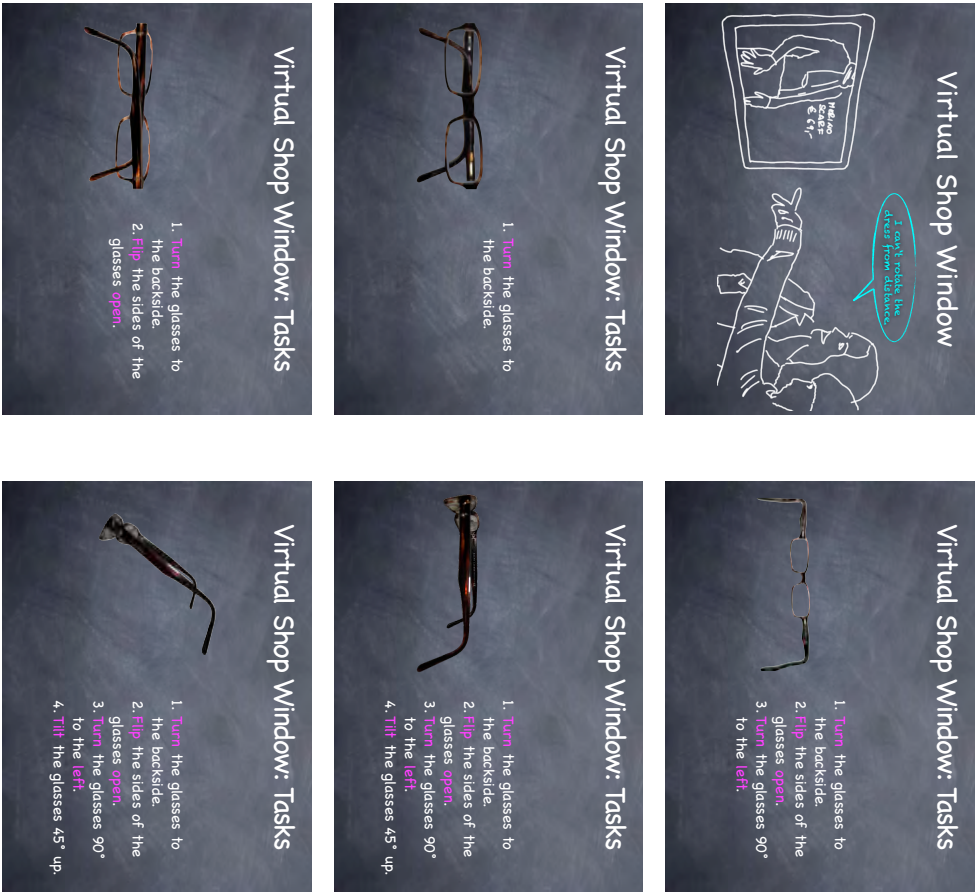


Figure A.5: A5: Interacting with a Virtual Shop Window

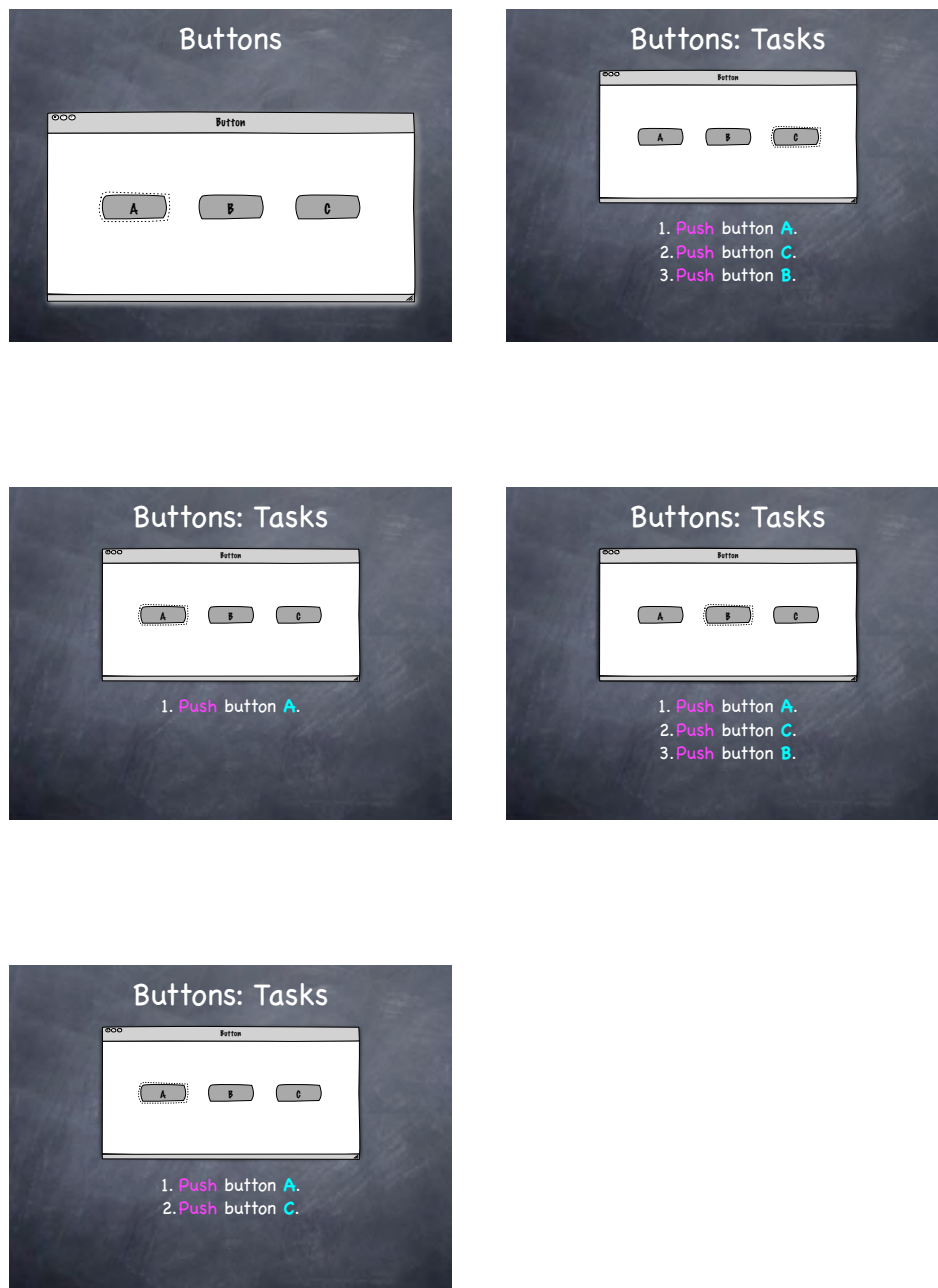
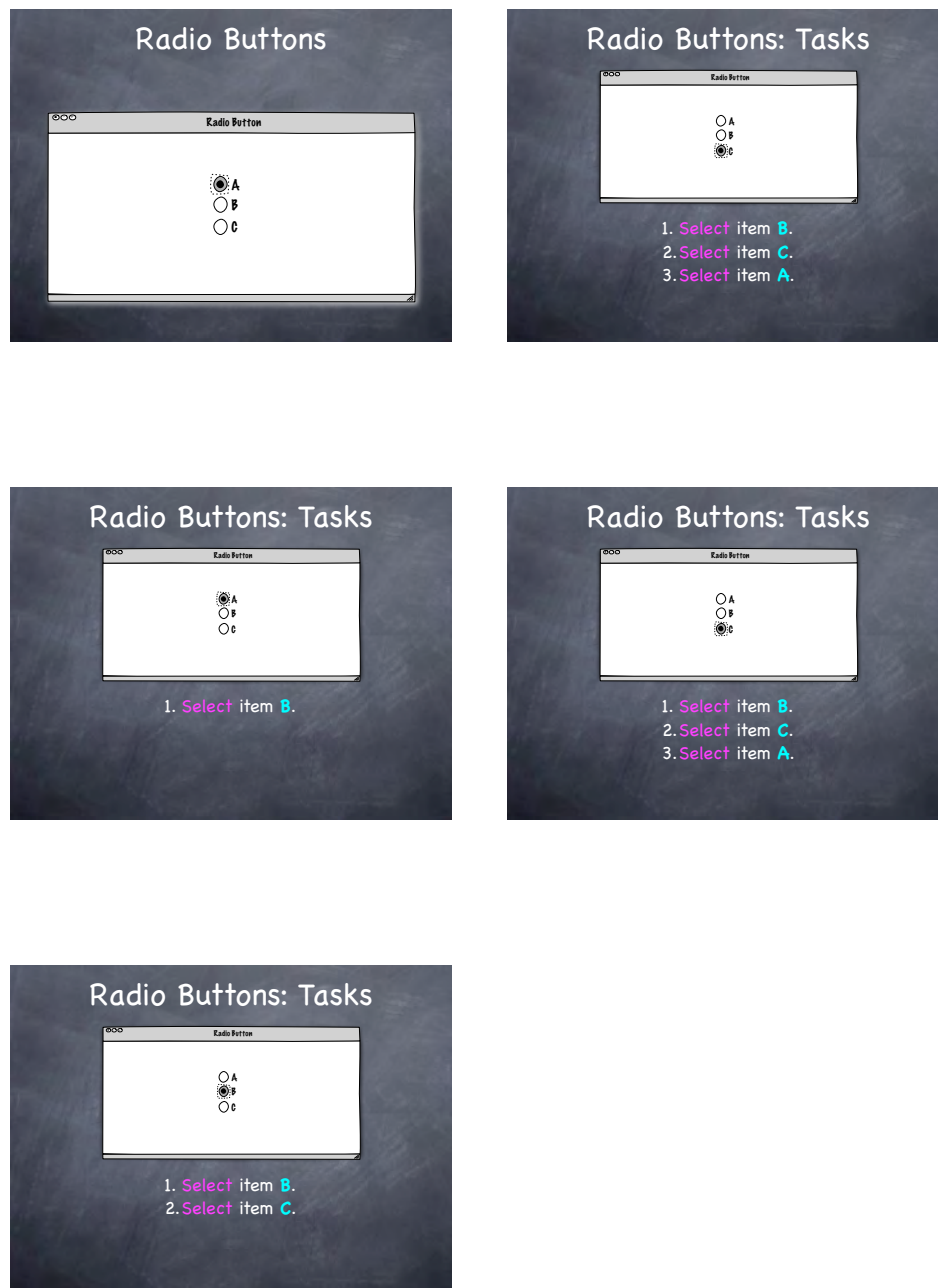


Figure A.6: B1: Buttons

**Figure A.7:** B2: *Radio Buttons*

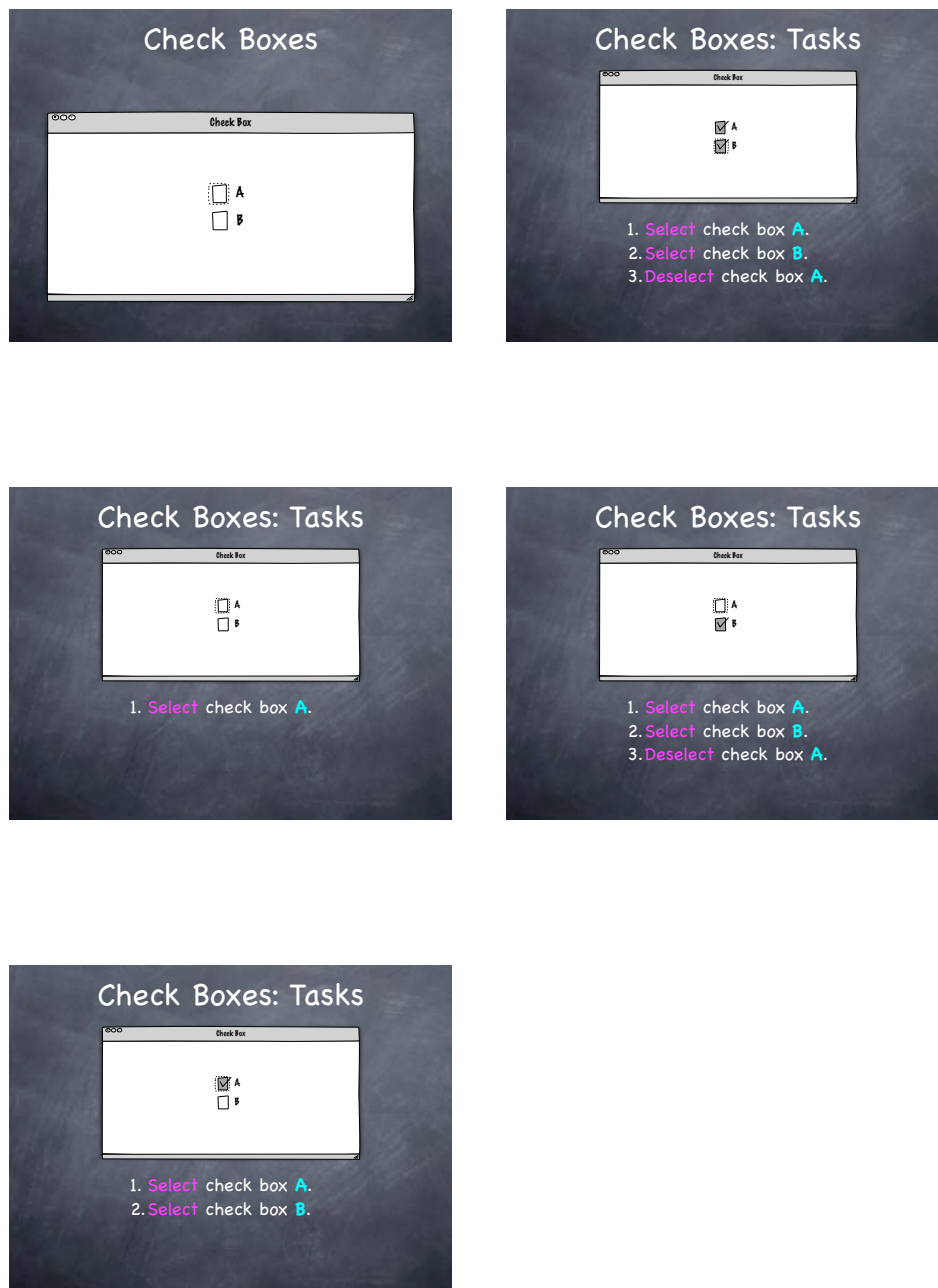
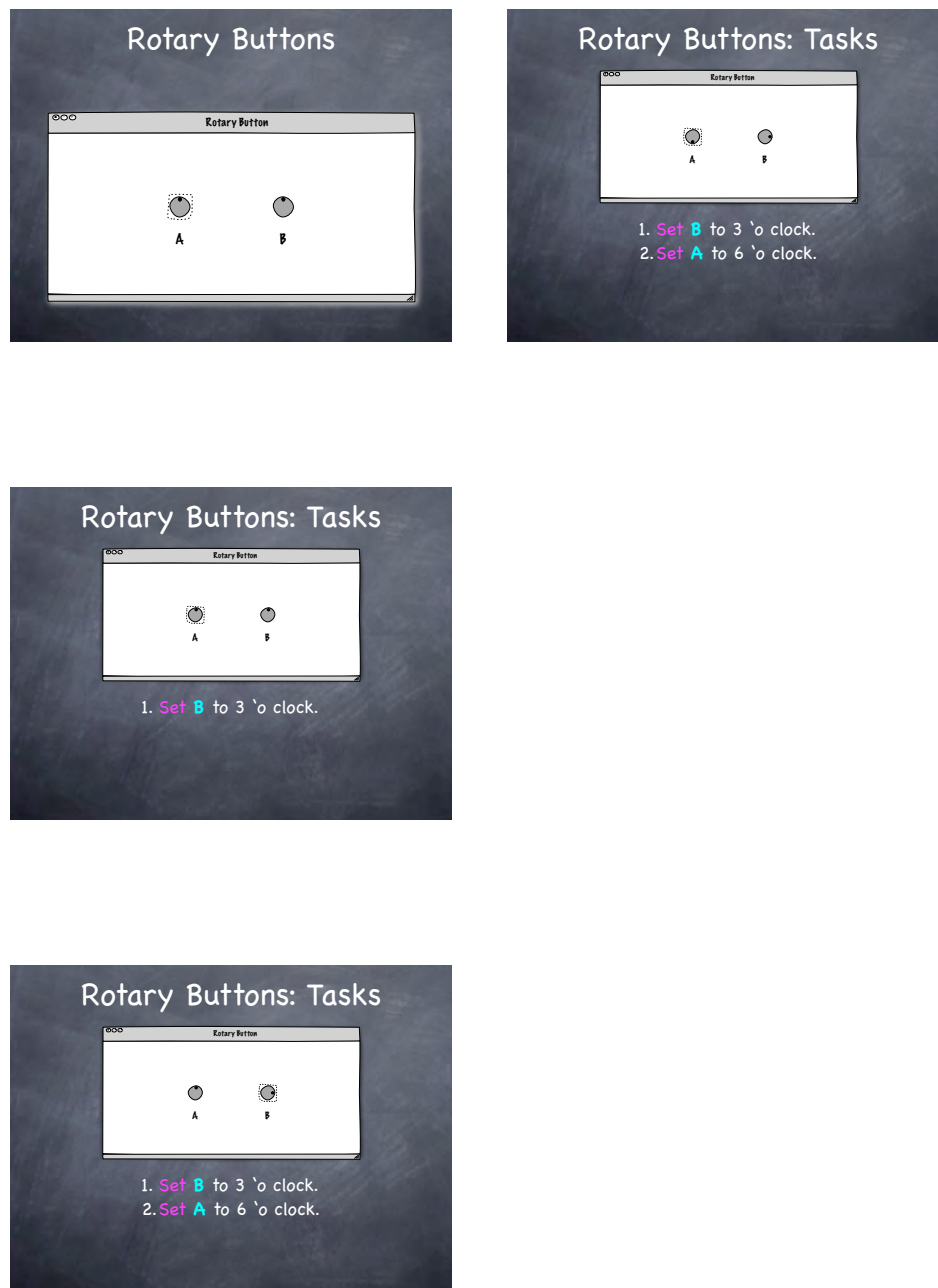


Figure A.8: B3: Check Boxes

**Figure A.9:** B4: *Rotary Buttons*

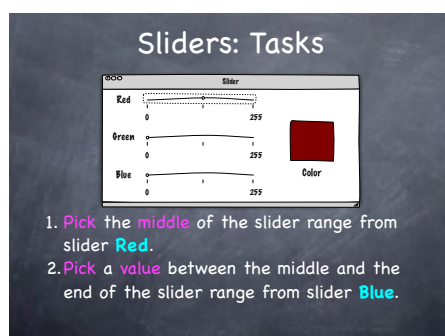
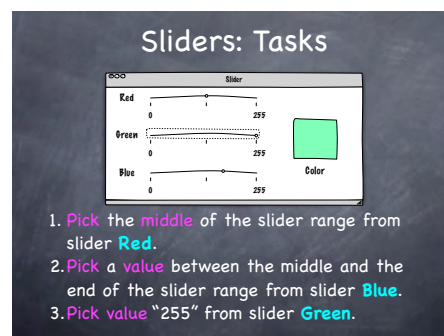
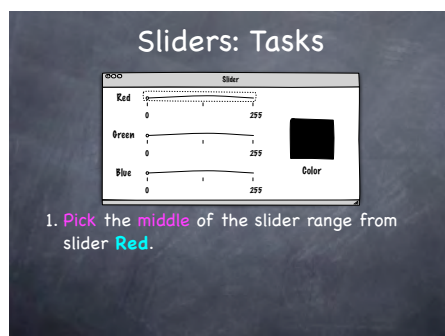
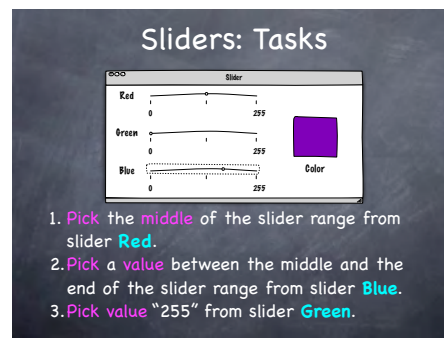
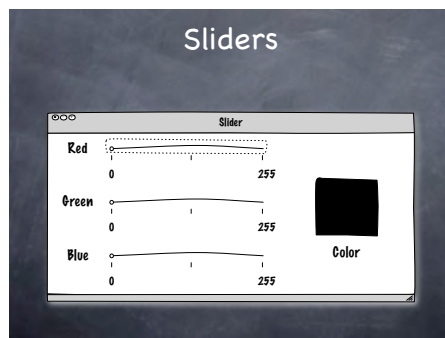


Figure A.10: B5: Sliders

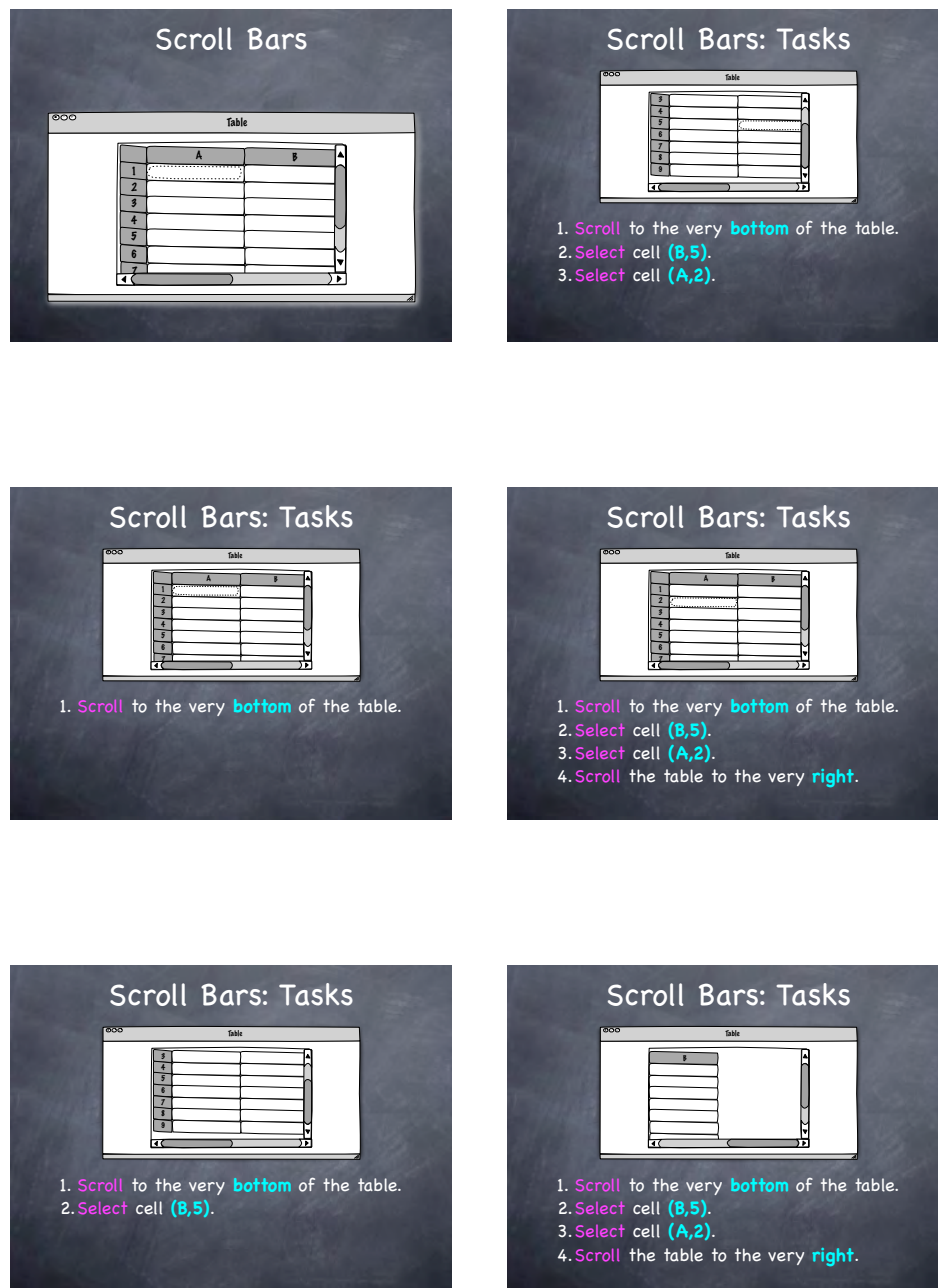


Figure A.11: B6: Scroll Bars

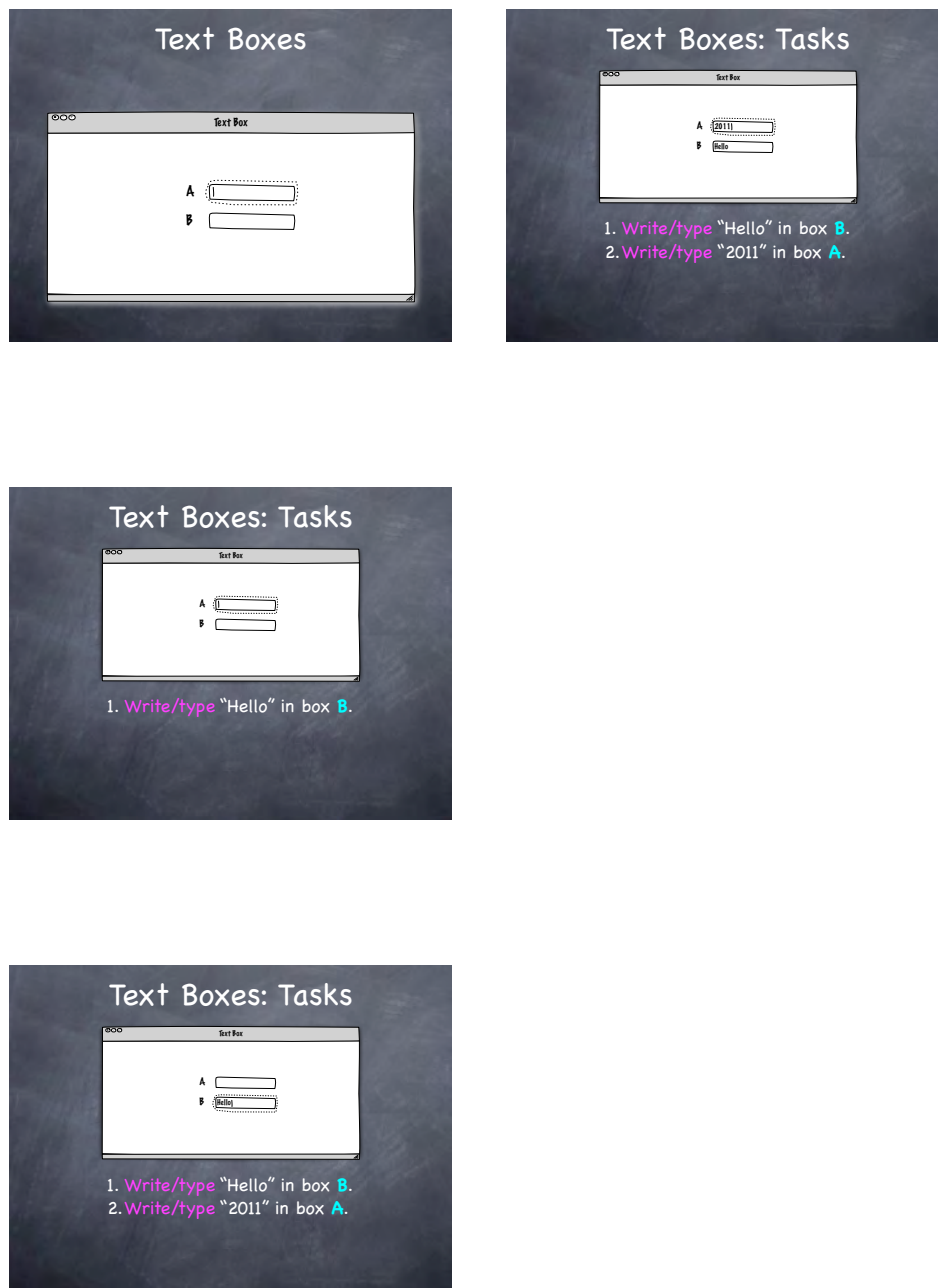


Figure A.12: B7: Text Boxes

Questions for both scenarios		Feedback Form for the Wizard of Oz Study				
1.	I find the idea about appropriating objects as input devices convincing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
		strongly disagree	disagree	neither	agree	strongly agree
	In case you find it <i>rather</i> not convincing: why?					
2.	I think it is problematic for the system to distinguish between normal object usage and appropriation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
		strongly disagree	disagree	neither	agree	strongly agree
	Do you have any ideas how the system could distinguish between normal usage and appropriation of the everyday objects?					
3.	To my mind, there were objects which were completely inadequate for solving any task					
	if so, please name the objects:					
4.	For solving the tasks, I preferred choosing objects which had/were					
(a)	<input type="radio"/> flexible/deformable	<input type="radio"/> rigid	<input type="radio"/> I had no preference			
(b)	<input type="radio"/> buttons/controls	<input type="radio"/> smooth	<input type="radio"/> I had no preference			
(c)	<input type="radio"/> lightweight	<input type="radio"/> heavy	<input type="radio"/> I had no preference			
(d)	<input type="radio"/> small	<input type="radio"/> large	<input type="radio"/> I had no preference			
5.	While using/appropriating objects for solving the tasks, I encountered unexpected behavior					
	if so, please explain what happened with the object(s):					

A – Remote Target Systems		Feedback Form for the Wizard of Oz Study				
6.	Finding objects for managing the tasks in the scenarios felt easy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
		strongly disagree	disagree	neither	agree	strongly agree
	presenting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	watching TV	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	playing a game	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	operating the light	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	virtual stop window	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7.	Finding gestures for appropriating objects for solving the tasks felt easy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
		strongly disagree	disagree	neither	agree	strongly agree
	presenting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	watching TV	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	playing a game	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	operating the light	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	virtual stop window	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8.	For managing the following tasks (as seen in the study), I expect to have a					
(a)	changing the volume	<input type="radio"/> discrete control	<input type="radio"/> continuous control	<input type="radio"/> I don't care		
(b)	muting the volume	<input type="radio"/> discrete control	<input type="radio"/> continuous control	<input type="radio"/> I don't care		
(c)	moving an avatar	<input type="radio"/> discrete control	<input type="radio"/> continuous control	<input type="radio"/> I don't care		
(d)	navigating in 3D	<input type="radio"/> discrete control	<input type="radio"/> continuous control	<input type="radio"/> I don't care		
(e)	dimming the light	<input type="radio"/> discrete control	<input type="radio"/> continuous control	<input type="radio"/> I don't care		
9.	I think that the objects I have chosen for each scenario are normally also available/in reach of that scenario (i.e., not only in a study but also in reality)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
		strongly disagree	disagree	neither	agree	strongly agree
	presenting	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	watching TV	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	playing a game	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	operating the light	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	virtual stop window	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure A.13: WOZ Feedback Form, pages 1 & 2

A – Remote Target Systems		Feedback Form for the Wizard of Oz Study				
10. For handling multiple similar tasks, I preferred using						
<div><input type="radio"/> multiple equal objects</div> <div><input type="radio"/> one multi-purpose object</div> <div><input type="radio"/> I don't know</div> <div><input type="radio"/> multiple various objects</div> <div><input type="radio"/> I don't care</div> <div><input type="radio"/> it depends:</div>						
11. While choosing objects, I carefully watched out for items with haptic properties						
<div><input type="radio"/> strongly disagree</div> <div><input type="radio"/> disagree</div> <div><input type="radio"/> neither</div> <div><input type="radio"/> agree</div> <div><input type="radio"/> strongly agree</div>						

B – GUI Scenarios		Feedback Form for the Wizard of Oz Study				
12. Finding objects for managing the tasks in the scenarios felt easy						
<div><input type="radio"/> strongly disagree</div> <div><input type="radio"/> disagree</div> <div><input type="radio"/> neither</div> <div><input type="radio"/> agree</div> <div><input type="radio"/> strongly agree</div>						
buttons						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
radio buttons						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
check boxes						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
rotary buttons						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
sliders						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
scroll bars						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
text boxes						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
13. Finding gestures (in general) for appropriating the objects for solving the tasks felt easy						
<div><input type="radio"/> strongly disagree</div> <div><input type="radio"/> disagree</div> <div><input type="radio"/> neither</div> <div><input type="radio"/> agree</div> <div><input type="radio"/> strongly agree</div>						
buttons						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
radio buttons						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
check boxes						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
rotary buttons						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
sliders						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
scroll bars						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						
text boxes						
<div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div> <div><input type="radio"/></div>						

B – GUI Scenarios		Feedback Form for the Wizard of Oz Study				
14. For managing the following tasks (as seen in the study), I expect to have a						
<div>(a) selecting a certain value with a slider</div> <div><input type="radio"/> discrete control</div> <div><input type="radio"/> continuous control</div> <div><input type="radio"/> I don't care</div>						
<div>(b) selecting a position with a rotary button</div> <div><input type="radio"/> discrete control</div> <div><input type="radio"/> continuous control</div> <div><input type="radio"/> I don't care</div>						
<div>(c) scrolling</div> <div><input type="radio"/> discrete control</div> <div><input type="radio"/> continuous control</div> <div><input type="radio"/> I don't care</div>						
<div>(d) typing</div> <div><input type="radio"/> discrete control</div> <div><input type="radio"/> continuous control</div> <div><input type="radio"/> I don't care</div>						
15. For handling multiple similar tasks, I preferred using						
<div><input type="radio"/> multiple equal objects</div> <div><input type="radio"/> one multi-purpose object</div> <div><input type="radio"/> I don't know</div> <div><input type="radio"/> multiple various objects</div> <div><input type="radio"/> I don't care</div> <div><input type="radio"/> it depends:</div>						
16. While choosing objects, I carefully watched out for items with haptic properties						
<div><input type="radio"/> strongly disagree</div> <div><input type="radio"/> disagree</div> <div><input type="radio"/> neither</div> <div><input type="radio"/> agree</div> <div><input type="radio"/> strongly agree</div>						
Any other feedback?						

Thank you for participating in this study!

Figure A.14: WOZ Feedback Form, pages 3 & 4

Appendix B

Appendix for the Interactive Prototype Study

Appendix B includes the feedback form for the interactive prototype study handed out to participants after they had completed both scenarios.

B.1 Feedback Form

In total, the feedback form included 17 questions.

Interactive Prototype User Study — Feedback Form

TV Scenario

1. I could handle the demanded tasks easily

	strongly disagree	disagree	neither	agree	strongly agree
(a) power on/off TV	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(b) go to a channel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(c) control the volume	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(d) control the brightness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. The interaction space was

	far too small	too small	decent	too big	far too big
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

3. To my mind, the predefined object-gesture-action mappings were intuitive/as I expected

	strongly disagree	disagree	neither	agree	strongly agree
(a) power the TV on/off by turning a cup	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(b) power the TV on/off by pushing the upper/lower third of a pack of pocket tissues	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(c) switch channels by pushing on the upper/lower third of a pack of pocket tissues	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(d) change the volume by turning a bottle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(e) confirm input by pushing the center of a pack of pocket tissues	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(f) cancel input by pushing on a wallet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4. Controlling a TV with everyday objects felt as comfortable as using a remote control

	far less comfortable	less comfortable	equally comfortable	more comfortable	far more comfortable
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. I could imagine to control a TV with everyday objects in case the dedicated input device (a remote control) is missing

	strongly disagree	disagree	neither	agree	strongly agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Interactive Prototype User Study — Feedback Form

Video Player Scenario

6. I could handle the demanded tasks easily

	strongly disagree	disagree	neither	agree	strongly agree
(a) play/pause the video	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(b) control the volume	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(c) control the temporal playback position	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(d) enter/exit fullscreen mode	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. The interaction space was

	far too small	too small	decent	too big	far too big
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

8. To my mind, the predefined object-gesture-action mappings were intuitive/as I expected

	strongly disagree	disagree	neither	agree	strongly agree
(a) play/pause the video by pushing a stapler	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(b) change the volume by turning a bottle	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(c) forward/rewind the video by sliding a glue stick	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
(d) toggle fullscreen mode by tapping a pack of pocket tissues	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. Navigating a movie with everyday objects felt as comfortable as using a mouse and/or keyboard

	far less comfortable	less comfortable	equally comfortable	more comfortable	far more comfortable
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. I could imagine to navigate a movie with everyday objects in case the dedicated input device (mouse and/or keyboard) is missing

	strongly disagree	disagree	neither	agree	strongly agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figure B.1: Interactive Prototype Feedback Form, pages 1 & 2

Interactive Prototype User Study — Feedback Form

General Questions

11. Principally, I needed eye contact with the objects during interaction

strongly disagree

disagree

neither

agree

strongly agree

12. Using an activation/deactivation gesture for being able to use the objects as controllers was annoying

strongly disagree

disagree

neither

agree

strongly agree

13. Pointing as activation/deactivation gesture felt natural

(a) activation gesture

strongly disagree

disagree

neither

agree

strongly agree

(b) deactivation gesture

strongly disagree

disagree

neither

agree

strongly agree

14. Hiding the object in order to deactivate it felt natural

(a) (pointing as activation gesture):

strongly disagree

disagree

neither

agree

strongly agree

(b) (pointing as deactivation gesture):

strongly disagree

disagree

neither

agree

strongly agree

15. As regards object deactivation, I preferred

pointing

strongly disagree

disagree

neither

agree

strongly agree

hiding

strongly disagree

disagree

neither

agree

strongly agree

no preference

strongly disagree

disagree

neither

agree

strongly agree

16. At any time, I was fully aware of which objects were activated and which objects were deactivated

(a) ... while using pointing for deactivation

strongly disagree

disagree

neither

agree

strongly agree

(b) ... while using hiding for deactivation

strongly disagree

disagree

neither

agree

strongly agree

17. I would have needed visual feedback indicating activated objects

strongly disagree

disagree

neither

agree

strongly agree

Figure B.2: Interactive Prototype Feedback Form, page 3

Bibliography

Eric Akaoka, Tim Ginn, and Roel Vertegaal. DisplayObjects: Prototyping Functional Physical Interfaces on 3D Styrofoam, Paper or Cardboard Models. In *TEI '10: Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*. ACM, January 2010.

Nao Akechi, Tsukasa Mizumata, and Ryuuki Sakamoto. Hovering Fingertips Detection on Diffused Surface Illumination. In *ITS '10: Proceedings of the ACM international conference on Interactive tabletops and surfaces*. ACM, November 2011.

Greg Borenstein. *Making Things See*. 3D Vision with Kinect, Processing, Arduino, and MakerBot. O'Reilly Media / Make, October 2011.

Bill Buxton. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007.

Andrew Carvey, Jim Gouldstone, Pallavi Vedurumudi, Adam Whiton, and Hiroshi Ishii. Rubber Shark as User Interface. *CHI EA '06: Proceedings of the 24th annual conference on Human factors in computing systems*, April 2006.

Kai-Yin Cheng, Rong-Hao Liang, Bing-Yu Chen, Rung-Huei Laing, and Sy-Yen Kuo. iCon: Utilizing Everyday Objects as Additional, Auxiliary and Instant Tabletop Controllers. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*. ACM, April 2010.

Keywon Chung, Michael Shilman, Chris Merrill, and Hiroshi Ishii. OnObject: Gestural Play with Tagged Everyday Objects. In *UIST '10: Adjunct proceedings of the 23rd*

- annual ACM symposium on User interface software and technology*. ACM, October 2010.
- Alan Dix, Janet E Finlay, Gregory D Abowd, and Russell Beale. *Human-Computer Interaction (3rd Edition)*. *Human-Computer Interaction (3rd Edition)*, December 2003.
- Kenneth P. Fishkin. A Taxonomy for and Analysis of Tangible Interfaces. *Personal and Ubiquitous Computing*, 8(5), September 2004.
- Jefferson Y Han. Low-cost Multi-touch Sensing through Frustrated Total Internal Reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*. ACM, October 2005.
- Chris Harrison, Desney Tan, and Dan Morris. Skinput: Appropriating the Body as an Input Surface. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*. ACM, April 2010.
- David Holman and Hrvoje Benko. SketchSpace: Designing Interactive Behaviors with Passive Materials. In *CHI EA '11: Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*. ACM, May 2011.
- David Holman and Roel Vertegaal. Organic User Interfaces: Designing Computers in any Way, Shape, or Form. *Communications of the ACM*, 51(6), June 2008.
- David Holman, Roel Vertegaal, Mark Altosaar, Nikolaus Troje, and Derek Johns. PaperWindows: Interaction Techniques for Digital Paper. In *CHI '05: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, April 2005.
- Wolfgang Hürst and Philipp Jarvers. Interactive, Dynamic Video Browsing with the Zoomslider Interface. In *ICME 2005: IEEE International Conference on Multimedia and Expo*, 2005.
- Hiroshi Ishii and Brygg Ullmer. Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, March 1997.

Shahram Izadi, Andrew Davison, Andrew Fitzgibbon, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, and Dustin Freeman. KinectFusion: Real-Time 3D Reconstruction and Interaction Using a Moving Depth Camera. In *UIST '11: Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, October 2011.

Martin Kaltenbrunner. reacTIVision and TUIO: A Tangible Tabletop Toolkit. In *ITS '09: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ACM, November 2009.

Boriana Koleva, Steve Benford, Kher Hui Ng, and Tom Rodden. A Framework for Tangible User Interfaces. Technical report, December 2011.

Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research Methods in Human-Computer Interaction*. John Wiley & Sons Inc., 2009.

Julian Lepinski and Roel Vertegaal. Cloth Displays: Interacting with Drapable Textile Screens. In *TEI '11: Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*. ACM, January 2011.

Pranav Mistry and Pattie Maes. Mouseless: A Computer Mouse as Small as Invisible. In *CHI EA '11: Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*. ACM, May 2011.

Pranav Mistry, Tsuyoshi Kuroki, and Chaochi Chang. TaPuMa: Tangible Public Map for Information Acquisition through the Things We Carry. In *Ambi-Sys '08: Proceedings of the 1st international conference on Ambient media and systems*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), February 2008.

Pranav Mistry, Pattie Maes, and Liyan Chang. WUW - Wear Ur World: A Wearable Gestural Interface. In *CHI '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*. ACM, April 2009.

- Jakob Nielsen and Robert L. Mack. *Usability Inspection Methods*. John Wiley & Sons, Inc., January 1994.
- Donald A. Norman. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Basic Books, 1993.
- Donald A. Norman. *The Design of Everyday Things*. Basic books, 2002.
- Jenny Preece, Yvonne Rogers, and Helen Sharp. *Interaction Design*. Beyond Human-Computer Interaction. Apogeo Editore, 2004.
- B. Ullmer and H. Ishii. Emerging Frameworks for Tangible User Interfaces. *IBM Systems Journal*, 39(3):915–931, 2000.
- Roel Vertegaal. The (Re)Usability of Everyday Computational Things. *interactions*, 18(1):38–41, January 2011.
- Mark Weiser. The Computer for the 21st Century. *Scientific American*, 265(3):94–104, 1991.
- Malte Weiss, Julie Wagner, Yvonne Jansen, Roger Jennings, Rasim Khoshabeh, James D. Hollan, and Jan Borchers. SLAP Widgets: Bridging the Gap Between Virtual and Physical Controls on Tabletops. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*. ACM, April 2009.
- Andrew D. Wilson and Hrvoje Benko. Combining Multiple Depth Cameras and Projectors for Interactions On, Above and Between Surfaces. In *UIST '10: Proceedings of the 23rd annual ACM symposium on User interface software and technology*. ACM, October 2010.
- Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. User-Defined Gestures for Surface Computing. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*. ACM, April 2009.
- Matthew Wright. Open Sound Control: An Enabling Technology for Musical Networking. *Organised Sound*, 10(3): 193–200, December 2005.
- Alexis Zerroug, Alvaro Cassinelli, and Masatoshi Ishikawa. Invoked Computing: Spatial Audio and Video AR Invoked through Miming. Technical report, March 2011.

Index

acceleration, 24
Apple iPhone, 107
Apple Keynote, 59–60
Apple QuickTime Player, 117
Apple ScriptingBridge Framework, 120
appropriation, 2, 170
AR, *see* Augmented Reality
Augmented Reality, 45

Cloth Displays, 32–34
Co-Objects, 25
contributions, 163

deactivation pattern, 152
deformation, 23
design space, 21
diary study, 51
Diffused Surface Illumination, 43, 107
direct manipulation, 21
DisplayObjects, 36–39
Distance Unit, 109
DSI, *see* Diffused Surface Illumination
DU, *see* Distance Unit
DualPlayer, 66

Event Processor, 112–113

fiducial, 111
Fiducial Controller, 112–113
form follows function, 33
Frustrated Total Internal Reflection, 43, 107
FTIR, *see* Frustrated Total Internal Reflection
future work, 163–165

Google, 1–2
Graphical User Interface, 64–65
GUI, *see* Graphical User Interface
Gulf of Execution, 7

HCI, *see* Human Computer Interaction

- HDTV, *see* High Definition Television
- Height Unit, 109
- hiding metaphor, 147
- High Definition Television, 118
- HU, *see* Height Unit
- Human Computer Interaction, 16

- iCon, 42–45, 105
- indirect manipulation, 22
- infrared, 107–108
- input equals output, 21–22, 33–34
- Instant User Interface, 6, 161
- Instant User Interface Guidelines, 167–171
- Inter Process Communication, 112–113
- Invoked Computing, 45–47
- iOS, 144
- IPC, *see* Inter Process Communication
- IR, *see* infrared
- IUI, *see* Instant User Interface

- KCPI, *see* Kinect Controller Programming Interface
- Kinect Controller, 112–113
- Kinect Controller Programming Interface, 113–114
- KinectFusion, 16

- LightSpace, 40

- mental model, 103, 130
- metaphorical compliance, 33
- Microsoft Kinect, 107
- Microsoft Xbox 360, 74
- MIDI, *see* Musical Instrument Digital Interface
- Mouseless, 26–28
- movement, 24
- Musical Instrument Digital Interface, 115

- natural mapping, 68
- non-stationary gesture, 24

- object competition, 154
- object preparation, 19–20
- object-gesture-action mapping, 57
- On-Screen Display, 118
- OnObject, 34–36
- Open Sound Control, 115
- Opportunistic Actions, 9
- Organic User Interface, 13
- OS X Lion, 117
- OSC, *see* Open Sound Control
- OSD, *see* On-Screen Display
- OUI, *see* Organic User Interface

- Physical Affordances, 2, 24–25, 168
Physical Instantiation, 69, 162, 170
PI, *see* Physical Instantiation
- qualitative counting, 52–53
- Radio Frequency Identification, 34–35
reactIVision, 43, 111
Red Green Blue, 107–108
remote target system, 64
RFID, *see* Radio Frequency Identification
RGB, *see* Red Green Blue
- Short Message Service, 53
SketchSpace, 39–42
Skinput, 28–29
SLAP Widgets, 11
SMS, *see* Short Message Service
Sony PlayStation, 74
stationary gesture, 23
- Tangible User Interface, 11
Tangible User Interface Object, 115–116
ThoMoNetworking, 117
touch, 23
TUI, *see* Tangible User Interface
TUIO, *see* Tangible User Interface Object
- UbiComp, *see* Ubiquitous Computing
Ubiquitous Computing, 14
UI widget, *see* User Interface widget
unintended deactivation, 149–150
User Interface widget, 64–65
- VGA, *see* Video Graphics Array
Vicon Motion Capture System, 32, 111–112
Video Graphics Array, 111
- Wacom Cintiq, 59–60
Wear Ur World, 30–32
WIMP, *see* Windows, Icons, Menus, Pointer
Windows, Icons, Menus, Pointer, 118
Wizard of Oz study, 57
WOZ study, *see* Wizard of Oz study
wrong object state, 149
WUW, *see* Wear Ur World

