



ROOPE RAISAMO

**Multimodal Human-
Computer Interaction:
a constructive and
empirical study**



ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty of Economics and Administration of the University of Tampere,
for public discussion in the Paavo Koli Auditorium of the University
on December 7th, 1999 at 12 o'clock.

*University of Tampere
Tampere 1999*



ROOPE RAISAMO

**Multimodal Human-
Computer Interaction:
a constructive and
empirical study**

*University of Tampere
Tampere 1999*

**Multimodal Human-
Computer Interaction:
a constructive and
empirical study**

ACADEMIC DISSERTATION

University of Tampere,
Department of Computer Science
Finland

ISBN 951-44-4702-6
ISSN 1456-954X

Contents

List of publications	iii
Acknowledgments.....	v
Abstract	vii
1. INTRODUCTION.....	1
2. MULTIMODAL INTERACTION.....	4
2.1. DEFINING MULTIMODAL INTERACTION	4
2.1.1. Multimodal interaction: a human-centered view	5
2.1.2. Multimodal interaction: a system-centered view.....	6
2.2. MODELING MULTIMODAL INTERACTION	9
2.2.1. Architecture of multimodal user interfaces	9
2.2.2. Fusion of input modalities.....	10
2.3. THE HISTORY OF MULTIMODAL USER INTERFACES.....	13
2.4. POTENTIAL BENEFITS OF MULTIMODAL INTERFACES	15
2.4.1. Natural and efficient dialogue	16
2.4.2. Mutual disambiguation of recognition errors	17
2.5. COMMON MISCONCEPTIONS ON MULTIMODAL INTERFACES.....	18
2.6. GUIDELINES FOR MULTIMODAL INTERACTION.....	20
2.7. SUMMARY	22
3. TWO-HANDED INTERACTION.....	24
3.1. THE PSYCHOLOGY OF TWO-HANDED INTERACTION	24
3.1.1. Analyzing manual tasks	24
3.1.2. Differences between the preferred and non-preferred hand	25
3.1.3. Division of labor between the hands	27
3.1.4. Cognitive advantages in two-handed interfaces.....	27
3.2. MODELING TWO-HANDED INTERACTION	28
3.3. THE HISTORY OF TWO-HANDED USER INTERFACES	32
3.4. POTENTIAL BENEFITS OF TWO-HANDED INTERFACES.....	34
3.4.1. Two hands are natural	34
3.4.2. Two hands are efficient.....	34
3.4.3. Two hands are less moded	35
3.5. GUIDELINES FOR TWO-HANDED INTERACTION.....	36
3.5.1. Independent interaction.....	36

3.5.2. Parallel interaction.....	37
3.5.3. Combined interaction.....	37
3.6. SUMMARY	38
4. MULTIMODAL INTERACTION TECHNIQUES	39
4.1. A NEW DIRECT MANIPULATION TECHNIQUE FOR ALIGNING OBJECTS IN DRAWING PROGRAMS	40
4.2. DESIGN AND EVALUATION OF THE ALIGNMENT STICK	41
4.3. AN ALTERNATIVE WAY OF DRAWING.....	42
4.4. AN EMPIRICAL STUDY OF HOW TO USE INPUT DEVICES TO CONTROL TOOLS IN DRAWING PROGRAMS	43
4.5. A MULTIMODAL USER INTERFACE FOR PUBLIC INFORMATION KIOSKS.....	44
4.6. EVALUATING TOUCH-BASED SELECTION TECHNIQUES IN A PUBLIC INFORMATION KIOSK	45
4.7. SUMMARY	46
5. FURTHER RESEARCH.....	48
5.1. ALIGNMENT TOOLS.....	48
5.1.1. Alternative shapes for alignment tools	49
5.1.2. Making alignment points and grouping more intuitive.....	49
5.1.3. Bridging the virtual and physical worlds	50
5.2. TWO-DIMENSIONAL SCULPTING METAPHOR.....	51
5.3. TOUCHSCREEN SELECTION TECHNIQUES	52
5.4. ENHANCED MULTIMODALITY	53
5.4.1. Feet in desktop interfaces	54
5.4.2. Machine vision techniques	54
5.5. SUMMARY	55
6. CONCLUSIONS	56
REFERENCES.....	58
 PAPER I	 69
PAPER II.....	79
PAPER III	105
PAPER IV	115
PAPER V	125
PAPER VI	133

List of publications

This thesis is based on the following research papers:

- I Roope Raisamo and Kari-Jouko Räihä, A new direct manipulation technique for aligning objects in drawing programs. *ACM UIST '96 Symposium on User Interface Software and Technology*, ACM Press, 1996, 157-164. [Raisamo and Räihä, 1996]

- II Roope Raisamo and Kari-Jouko Räihä, Design and evaluation of the alignment stick. *Interacting with computers* (accepted for publication). [Raisamo and Räihä, 1999]

- III Roope Raisamo, An alternative way of drawing. *Human Factors in Computing Systems, CHI '99 Conference Proceedings*, ACM Press, 1999, 175-182. [Raisamo, 1999a]

- IV Roope Raisamo, An empirical study of how to use input devices to control tools in drawing programs. *OzCHI '99 Conference Proceedings, 1999 Conference of the Computer Human Interaction Special Interest Group of the Ergonomics Society of Australia*, School of Information Studies, Charles Sturt University, 1999, 71-77. [Raisamo, 1999b]

- V Roope Raisamo, A multimodal user interface for public information kiosks. *Proceedings of 1998 Workshop on Perceptual User Interfaces (PUI '98)*, Matthew Turk (Ed.), San Francisco, November 1998, 7-12. [Raisamo, 1998]

- VI Roope Raisamo, Evaluating different touch-based interaction techniques in a public information kiosk. Report A-1999-11, Department of Computer Science, University of Tampere, 1999. [Raisamo, 1999c]

Acknowledgments

I have now reached the first major goal in my research career. This dissertation took four years of work and required a great amount of digging into the literature, and designing and programming of the two prototype systems. Great many hours were also spent in carrying out the four empirical studies that evaluated the systems and new interaction techniques. I would like to take the opportunity to thank all the people who have supported me during my research and writing process.

First, I want to thank my supervisor, professor Kari-Jouko Räihä. Without his continuous support this dissertation might not even exist. Starting with my part-time research position in the research project “User interfaces” in the spring of 1995, he has provided me with excellent research conditions. He also helped in my selection as one of the first researchers who were financed by Tampere Graduate School in Information Science and Engineering (TISE) that granted me a four-year financing period and job security. Without that support I might have been seduced to make use of one of the industrial working positions that I have been offered within the past few years.

My four years were not limited to sitting in a small room in Tampere. I had the opportunity to visit many international conferences and work two summers in two prestigious research laboratories in the United States. Working in the heart of the Silicon Valley in the summer of 1996 for Digital Equipment Corporation’s Systems Research Center (Palo Alto, California) was great and gave me great insight in the work done by Digital and its rivals. Especially, I want to thank my host Dr. Marc H. Brown for allowing me to work with an interesting research project and learn more Java in the process.

The following summer I was selected to work as a summer intern in Digital Equipment Corporation’s Cambridge Research Laboratory (Cambridge, Massachusetts). It was a motivating research environment, and I want to thank my host Dr. Andrew Christian for suggesting an interesting research project that experimented with using speech recognition in public information kiosks.

The Department of Computer Science in the University of Tampere has been a great place to do research, and the heads of the department, professor Räihä followed by professor Järvinen and professor Visala, have been cooperative in financing my research and related expenses. The administrative staff, Marja

Liisa Nurmi, Tuula Moisio and Heli Rikala have helped me with the bureaucracy and deserve thanks for that. Many colleagues in the department have supported me in my work, especially the people in the TAUCHI Group (Tampere University Computer-Human Interaction Group). Participating in the interdisciplinary project “Multimodal human information processing and human-computer interaction” has given me a wider view of multimodality and introduced me to some psychological research. For that I wish to thank the project leader, professor Mikko Sams, who arranged many seminars in the project.

We have been lucky to have many great visitors from foreign universities and research laboratories. I want to thank the following people for their insights and suggestions concerning my work: Stephen Brewster (University of Glasgow, UK), Tom Hewett (Drexel University, USA), Kevin Mullet (Netscape, USA), Scott MacKenzie (York University, Canada) and Marilyn Tremaine (Drexel University, USA). I especially thank professor Scott MacKenzie for being willing to act as my opponent in the public defence of this dissertation.

I also thank the reviewers of the dissertation, professor Mikko Sams and professor Martti Mäntylä from the Helsinki University of Technology, for their comments and for being able to meet the tight deadline that we suggested.

Empirical tests require a lot of work and voluntary users. I want to thank Antti Aaltonen and Toni Pakkanen for helping me with carrying out two of the user studies. In addition, I want to express my gratitude to all those students who volunteered to test my prototypes in my course “Introduction to Computing”. The same applies to those colleagues who participated in some of the tests and to the students of English philology and human-computer interaction who participated in the last study.

For financing my research I want to thank the Academy of Finland, the Tampere Graduate School in Information Science and Engineering (TISE), the Department of Computer Science in the University of Tampere, and Digital Equipment Corporation.

Last but not least, I wish to express my gratitude to my parents and grandparents who have supported and encouraged me throughout my studies. My other relatives have also been very supportive. Moreover, it has been very important to have friends with whom to get away from the daily routines.

Tampere, November 1999

Roope Raisamo

Abstract

Multimodal interaction is a way to make user interfaces natural and efficient with parallel and synergistic use of two or more input or output modalities. Two-handed interaction is a special case of multimodal interaction that makes use of both hands in a combined and coordinated manner. This dissertation gives a comprehensive survey on issues that are related to multimodal and two-handed interaction. Earlier work in human-computer interaction and related psychology is introduced within both these fields.

The constructive part of this dissertation consists of designing and building a group of multimodal interaction techniques that were implemented in two research prototypes. The first prototype is an object-oriented drawing program that implements new tools that are controlled with two-handed input. The second prototype is a multimodal information kiosk that responds to both touch and speech input, and makes use of touch pressure sensing.

In order to evaluate the success of constructive research, four empirical studies were conducted to compare the new interaction techniques to the conventional methods and to evaluate how the users react on them. The first of the studies compared a new direct manipulation tool to conventional menu and palette commands. The second evaluation was more informal and determined how an alternative way of drawing would be used by normal users. The third study was aimed at determining what is the best input device configuration to control the new tools. The last study evaluated different touch-based selection techniques in a multimodal touch and speech based information kiosk.

The need for extensive interdisciplinary research is pointed out with a group of research questions that need to be answered to better understand multimodal human-computer interaction. The current knowledge only applies to a few special cases, and there is no unified modality theory of multimodal interaction that covers both input and output modalities.

CR Categories and Subject Descriptors:

- H.5.2 Information Systems - Information Interfaces and Presentation - User interfaces: *Interaction styles, Input devices and strategies*
- I.3.4 Computing Methodologies - Computer graphics - Graphics utilities: *Graphics editors*
- I.3.6 Computing Methodologies - Computer graphics - Methodology and techniques: *Interaction techniques*
- D.2.2 Software - Software engineering - Tools and techniques: *User interfaces*

1. Introduction

Human-computer interaction is a field that has developed rapidly in the last decade. Traditionally, most of the resources of the computer industry have been directed to developing faster microprocessors and high-capacity storage devices. For a long time the user was neglected when developing computer hardware and software. Fortunately, this has changed. Many researchers agree with Donald A. Norman [1990, p. 210]:

“The real problem with the interface is that it is an interface. Interfaces get in the way. I don’t want to focus my energies on an interface. I want to focus on the job.”

A common way to state the goal of getting rid of an interface is to call new interface prototypes “natural”. The term *natural user interface* is not an exact expression, but usually means an interface that is as easy to use and seamless as possible. The feeling of using an interface can be faded out by not attaching any interaction devices to the user and by designing the dialogue in a way that really is natural and understandable for the user.

Graphical user interfaces that emerged in the beginning of 1980s were a great leap towards user-friendly computing. They provided the user with a common look and feel, visual representations of data, and direct control using a new device called mouse together with the older keyboard. Personal computers were the vehicle for the spreading of these new user interfaces. Apple MacOS, Microsoft Windows and the X Window System are well-known graphical user interfaces that have developed slightly better all the time since the introduction of Xerox Star workstations [Lipkie *et al.*, 1982; Bewley *et al.*, 1983; Baecker and Buxton, 1987] and Apple Lisa [Baecker and Buxton, 1987; Perkins *et al.*, 1997].

The phrase “direct manipulation” was introduced by Ben Shneiderman [1982; 1983]. According to his definition, a direct manipulation interface should:

- present the objects visually with an understandable metaphor;
- have rapid, complementary and reversible commands;
- display the results of an action immediately; and
- replace writing with pointing and selecting.

Graphical user interfaces are based on Shneiderman’s principles, and many people consider mouse-based direct manipulation as the ultimate goal of user

interfaces. Naturally, other views exist. Among others, Pattie Maes has many times pointed out some deficiencies of direct manipulation. She bases her position on the fact that when more and more people are going to use computers, the machines should be active and not passively wait for the user to carry out tasks manually using direct manipulation. Maes' solution [Miller, 1997; Shneiderman and Maes, 1997] is to use software agents that actively observe the user and carry out routine tasks.

Computer hardware has always been an important factor that has an effect on all software development. User interfaces are not an exception. Graphical user interfaces became possible to implement when computer hardware could produce accurate bitmap displays and there was enough processing power to interactively manipulate such accurate screen presentations. These interfaces have developed smoother and their fine-tuning is still in progress. However, current hardware is much more advanced than the one with which graphical user interfaces were first implemented. We should be able to come up with a similar kind of revolution in user interfaces that graphical user interfaces caused in 1980s.

Multimodal user interfaces are a strong candidate for being the next breakthrough in building better user interfaces. Multimodal interfaces combine many simultaneous input modalities and may present the information using synergistic representation of many different output modalities. The input modalities can be as simple as two pointing devices, or they may include advanced perception techniques like speech recognition and machine vision. The simple cases do not require more processing power than the current graphical user interfaces, but still provide the user with more degrees of freedom with two continuous input feeds. New perception technologies that are based on speech, vision, electrical impulses or gestures usually require much more processing power than the current user interfaces.

Two-handed interfaces are a special case of multimodal user interfaces. The need for processing power depends on the kind of two-handed application that is implemented. Even a simple case, using two pointing devices in a normal graphical user interface, has been found to be more efficient and understandable than the basic mouse and keyboard interfaces [Buxton and Myers, 1986; Kabbash *et al.*, 1994; Zhai *et al.*, 1997]. Two-handed interfaces have also been found to make three-dimensional virtual environments better understood and more easily manipulated [Hinckley *et al.*, 1997a; 1997b; 1998a].

Of course, multimodal user interfaces and two-handed interfaces need to be designed well in order to benefit from them. It has been found that when these

interfaces are designed poorly they are neither better understood nor efficient [Dillon *et al.*, 1990; Kabbash *et al.*, 1994]. It is very important that human abilities are carefully analyzed and understood when these new interfaces are designed. Thus, careful consideration of human cognitive abilities and motor skills must be carried out. In addition, empirical tests are necessary to find out how new interaction ideas perform in reality. To be able to test the ideas empirically we need to do constructive research to implement research prototypes that implement the new ideas. This is why I have adopted a constructive and empirical perspective in this work.

In this dissertation I present a collection of multimodal interaction techniques. The research papers explain the details of the new interaction techniques. There are three cases of multimodal interaction techniques. These are selected examples of possibilities that alternative interaction techniques can offer to us. The techniques were implemented in two research prototypes, an object-oriented drawing program *R2-Draw* and a multimodal information kiosk framework *Touch'n'Speak*.

The first technique is based on two-handed interaction and presents a new tool to align objects in drawing programs and diagram editors. The tool was designed to be more intuitive and efficient than the conventional command-based alignment menus and palettes.

The second group of techniques follows from a generalization of the first two-handed tool and introduces an alternative interaction style for object-oriented drawing programs. The emphasis in these two cases was in defining new direct manipulation tools that can replace difficult commands or less direct ways of performing certain tasks.

The third case of using multimodal interaction techniques is a multimodal information kiosk that is based on touchscreen input and speech recognition. There the emphasis was in developing different selection techniques that are based on time, pressure and direct manipulation. A conversational user interface with speech commands and speech synthesis was used together with these touch techniques.

This dissertation provides background for the papers and a literature survey of multimodal and two-handed interaction. In chapter 2, I begin with the definition and analysis of multimodal interaction. Next, in chapter 3, two-handed interaction is discussed as a special case of multimodal interaction. Chapter 4 introduces the research papers. Chapter 5 discusses the practical consequences of the presented work and suggests future directions for the research. Chapter 6 summarizes the work.

2. Multimodal interaction

Despite its many fine qualities a graphical user interface makes quite poor use of human abilities. Buxton [1986] describes what a future anthropologist might conclude if he or she found a fully stocked computer store with all the equipment and software in working order:

“My best guess is that we would be pictured as having a well-developed eye, a long right arm, a small left arm, uniform-length fingers and a ‘low-fi’ ear. But the dominating characteristics would be the prevalence of our visual system over our poorly developed manual dexterity.”

This example demonstrates many current deficiencies in human-computer interfaces. If we compare using a computer to driving a car, we can see how badly the current computing systems make use of the human sensory and motor systems.

The research on multimodal user interfaces aims at making Buxton’s anthropologist example obsolete. The goals behind multimodal user interfaces are to make use of many different abilities that humans have. The emphasis is in combining different input or output channels in order to make using a computer more efficient, easier or both. Another way to describe multimodal interaction is that it can be used to make direct manipulation more direct.

2.1. Defining multimodal interaction

There are two views on multimodal interaction. The first one is heavily rooted in psychology and focuses on the human side: perception and control. There the word *modality* refers to human input and output channels. The second view is that of the most computer scientists and focuses on using two or more computer input or output modalities to build systems that make synergistic use of parallel input or output of these modalities. These two views are discussed in this section in an attempt to define multimodal interaction. Figure 1 presents a model for the identification of basic processes in multimodal human-computer interaction and comprises both views [MIAMI, 1995, p. 2].

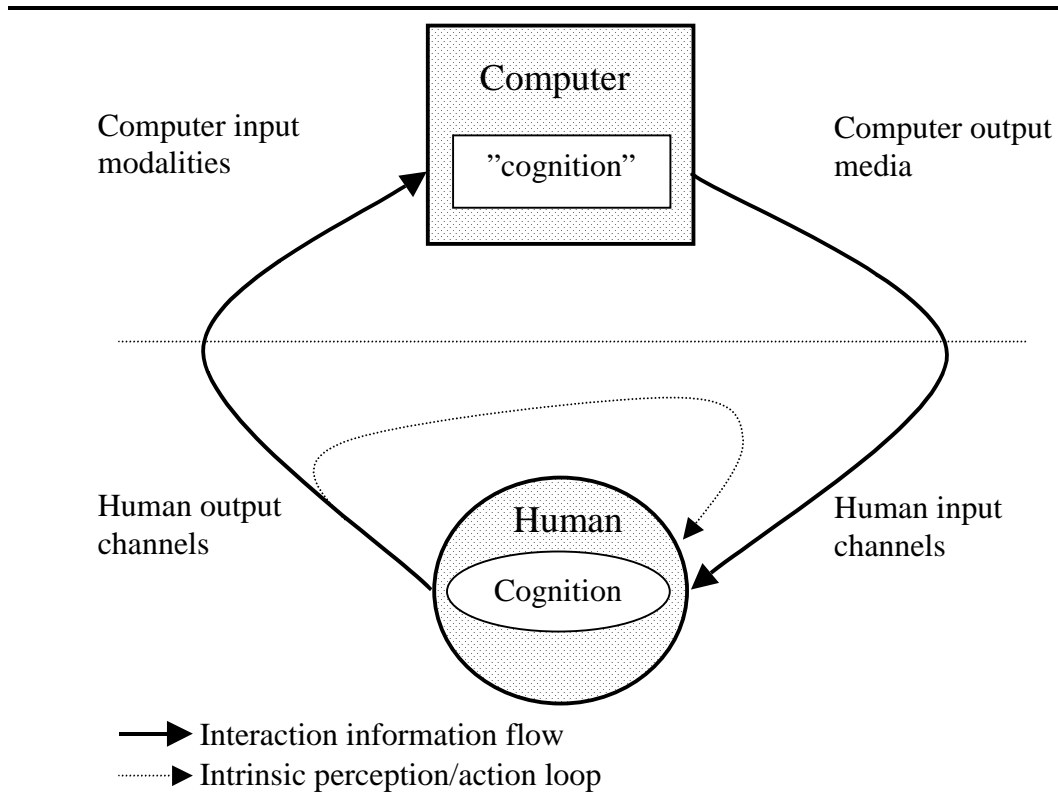


Figure 1. A model for the identification of basic processes in human-computer interaction [MIAMI, 1995].

2.1.1. Multimodal interaction: a human-centered view

In the human-centered view of multimodal interaction the focus is on *multimodal perception and control*, that is, human input and output channels. *Perception* means the process of transforming sensory information to higher-level representations [MIAMI, 1995, p. 15]. A *communication channel* consists of the sense organs, the nerve tracts, the cerebrum, and the muscles [Charwat, 1992]. A great deal of research has been carried out on the process of multimodal audio-visual perception [McGurk and MacDonald, 1976; MacDonald and McGurk, 1978; Sams *et al.*, 1998]. A talking head, three-dimensional animated human-like figure using speech synthesis [Parke and Waters, 1996], is an artifact that combines both visual and auditory computer outputs to create a multimodal audio-visual stimulus for the user.

In this view the concept of modality is tightly coupled with human senses. Silbernagel [1979] lists the different senses and their corresponding modalities that are shown in Figure 2. The figure shows a somewhat simplified version of perception. For example, tactile feedback is perceived through nerve endings that reside in deep tissue in addition to nerve terminals that are in the skin.

Sensory perception	Sense organ	Modality
Sense of sight	Eyes	Visual
Sense of hearing	Ears	Auditive
Sense of touch	Skin	Tactile
Sense of smell	Nose	Olfactory
Sense of taste	Tongue	Gustatory
Sense of balance	Organ of equilibrium	Vestibular

Figure 2. Different senses and their corresponding modalities [Silbernagel, 1979].

If we observe the modalities from a neurobiological point of view [Kandel and Schwartz, 1981; Shepherd, 1988], we can divide the modalities in seven groups: *internal chemical* (blood oxygen, glucose, pH), *external chemical* (taste, smell), *somatic senses* (touch, pressure, temperature, pain), *muscle sense* (stretch, tension, joint position), *sense of balance*, *hearing*, and *vision*. Since internal chemical senses are not obviously applicable in seamless user interfaces, we can safely ignore them in this context. The other neurobiological modalities are included in the modalities presented in Figure 2, but muscle sense (kinesthesia) can not be found in the figure. Kinesthesia is very important in determining where our hands and other body parts are at a given moment.

Sometimes we want to distinguish between *bimodal* and *multimodal* use of modalities. The word *bimodal* is used to nominate multimodal perception or control that makes use of exactly two modalities.

2.1.2. Multimodal interaction: a system-centered view

In computer science multimodal user interfaces have been defined in many ways. Chatty [1994] gives a summary of definitions for multimodal interaction by explaining that most authors who recently wrote on that subject considered it as defining systems that feature multiple input devices (multi-sensor interaction) or multiple interpretations of input issued through a single device.

Chatty's explanation of multimodal interaction is the one that most computer scientists use. With the term *multimodal user interface* they mean a system that accepts many different inputs that are combined in a meaningful way. For a long time computer output developed much faster than the ways that we can use to control the systems. If we compare the current multimedia presentation capabilities of the computers to the input devices that we use to

control these machines, there is a great difference in the sophistication and diversity of available input and output channels.

Nigay and Coutaz [1993] define multimodality in the following way:

“*Multimodality* is the capacity of the system to communicate with a user along different types of communication channels and to extract and convey meaning automatically.”

Both multimedia and multimodal systems use multiple communication channels. Nigay and Coutaz differentiate between these systems by stating that a multimodal system is able to automatically model the content of the information at a high level of abstraction. A *multimodal system strives for meaning*. For example, an electronic mail system that supports voice and video clips is not multimodal if it only transfers them to another person and does not interpret the inputs.

According to Nigay and Coutaz the two main features of multimodal interfaces are the following:

- the fusion of different types of data from/to different input or output devices, and
- the temporal constraints imposed on information processing from/to input or output devices.

They have used these features to define a design space for multimodal systems that is presented in Figure 3.

		Use of modalities			
		Sequential		Parallel	
Fusion	Combined	ALTERNATE		SYNERGISTIC	
	Independent	EXCLUSIVE		CONCURRENT	
		Meaning	No Meaning	Meaning	No Meaning
		Levels of abstraction			

Figure 3. Design space for multimodal systems [Nigay and Coutaz, 1993].

The different dimensions of the design space define eight possible classes of systems. By definition [Nigay and Coutaz, 1993], multimodal systems require the value “Meaning” in the “Levels of abstraction”. Thus, there are four distinct classes of multimodal systems: exclusive, alternate, concurrent and synergistic.

“Use of modalities” refers to the temporal dimension. Parallel operation can be achieved at different levels of abstraction [Coutaz *et al.*, 1993]. The most important level is the task level, the level that the user interacts with. It must

seem to be concurrent in order to make the user perceive the system as parallel. Low-level, or physical level, concurrency is not a requirement for a multimodal system, but needs to be considered in the implementation. Most current computing systems are able to offer an imitation of concurrent processing even though the processing is in fact sequential.

Fusion is the most demanding criterion in the design space. Here “Combined” means that different modalities are combined in synergistic higher-level input tokens. “Independent” means that there are parallel but not linked modalities in the interface. The synergistic use of modalities implies the fusion of data from different modeling techniques. Nigay and Coutaz [1993] have identified three levels of fusion: lexical, syntactic and semantic. They can be mapped to the three conceptual design levels defined by Foley *et al.* [1990, pp. 394-395]:

- *Lexical fusion* corresponds to conceptual Binding level. It happens when hardware primitives are bound to software events. An example of lexical fusion is selecting multiple objects when the shift key is down.
- *Syntactic fusion* corresponds to conceptual Sequencing level. It involves the combination of data to obtain a complete command. Sequencing of events is important at this level. An example of syntactic fusion is synchronizing speech and pen input in a map selection task.
- *Semantic fusion* corresponds to conceptual Functional level. It specifies the detailed functionality of the interface: what information is needed for each operation on the object, how to handle the errors, and what the results of an operation are. Semantic fusion defines meanings but not the sequence of actions or the devices with which the actions are conducted. An example of semantic fusion is a flight route selection task that requires at least two airports as its input through either touch or speech and draws the route on a map.

Still, having defined that multimodal systems make use of multiple input or output modalities does not define the properties of the actual systems very well. There are two main categories of systems that have very different underlying paradigms:

1. Multimodal input modalities are used to enhance direct manipulation behavior of the system. The machine is a passive tool and tries to understand the user through all different input modalities that the system recognizes. The user is always responsible for initiating the operations. Thus, if the user does not know what to do, nothing gets

done. The goal is to use *the computer as a tool* and it follows Ben Shneiderman's [1982; 1983] principles of direct manipulation interfaces.

2. The multiple modalities are used to increase anthropomorphism in the user interface: *the computer as a dialogue partner*. This paradigm makes multimodal output important, and the systems in this category often make use of talking heads, speech recognition and other human-like modalities. This kind of multimodal system can often be described as an agent-based conversational user interface.

The selection between these two main categories or other types of systems must be done on the basis of specific system requirements. These categories are not exclusive: A tool-like multimodal user interface may have a tutor that is presented as a dialogue partner, whereas a conversational user interface may have multimodal interaction techniques that directly respond to the user's manual commands.

In this dissertation, I have followed the system-centered view on multimodal user interfaces and focus on multimodal input which takes into account psychological theories of human motor-sensory control.

2.2. Modeling multimodal interaction

There are several problems in designing a multimodal user interface. First, the set of input modalities must be selected right. This is no trivial problem and depends greatly on the task the interface will be used for. This dissertation aims at answering this question in some classes of user interfaces. Another, even greater problem is how to combine different input channels. First, I will discuss multimodal systems in general. A high-level model of multimodal human-computer interaction was already presented in Figure 1. Here we take a look at a more detailed model by Maybury and Wahlster [1998, p. 3]. Some models for the fusion of input modalities have been developed by Nigay and Coutaz [1993; 1995] and Moran *et al.* [1997]. These models will be described later in this section.

2.2.1. Architecture of multimodal user interfaces

Maybury and Wahlster [1998, p. 3] describe a high-level architecture of intelligent user interfaces. As their notion of intelligent user interfaces includes multimodal interaction, this model can be used for modeling multimodal interfaces. Figure 4 shows an adaptation of their model that presents the different levels of processes that are contained in the complete architecture.

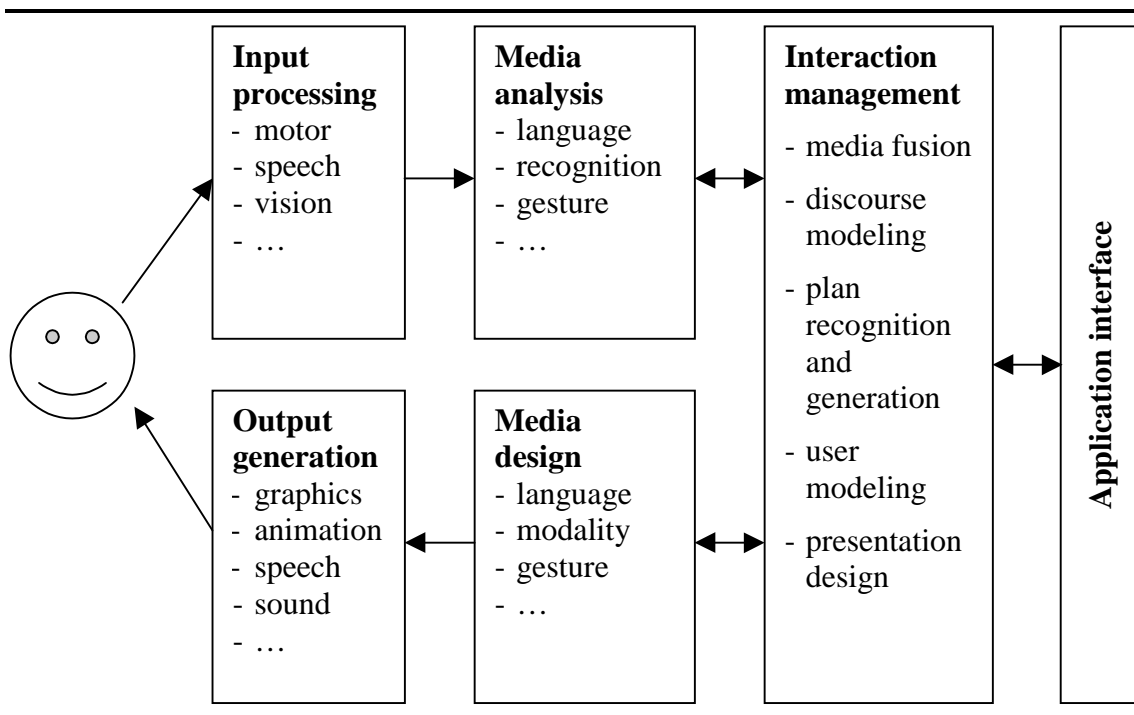


Figure 4. An architecture of multimodal user interfaces.
Adapted from [Maybury and Wahlster, 1998].

The presented architecture contains many models for different processes in the system. Each of these models can be refined to fulfill the requirements of a given system. Specifically, user and discourse models are highly important in a multimodal interface. There can be one or more user models in a given system. If there are several user models, the system can also be described as an adaptable user interface. The discourse model handles the user interaction in a high level, and uses media analysis and media design processes to understand what the user wants and to present the information with the appropriate output channels. Wahlster [1991] discusses user and discourse models for multimodal communication when he describes an intelligent multimodal interface to expert systems.

Here I focus on the fusion of input modalities that is a central property of synergistic multimodal user interfaces. Next, I will discuss two examples of fusion architectures.

2.2.2. Fusion of input modalities

A multiagent architecture seems to be appropriate for building complex multimodal interfaces and has been used in both of the systems that are presented next. This architecture structures the input nicely by assigning one or

more agents to each input modality and by providing dialogue control and fusion as separate agents in the system. A multiagent system inherently supports parallel processing and may also provide means for distributed computing.

PAC-Amodeus

Nigay and Coutaz [1993; 1995] present a fusion mechanism that is used with their PAC-Amodeus software architecture model. They use a melting pot metaphor to model a multimodal input event. PAC agents [Coutaz, 1987] act in this multiagent system and are responsible for handling these events. Figure 5 shows how the melting pots are composed from multiple inputs in a system similar to Bolt's [1980] Put-That-There.

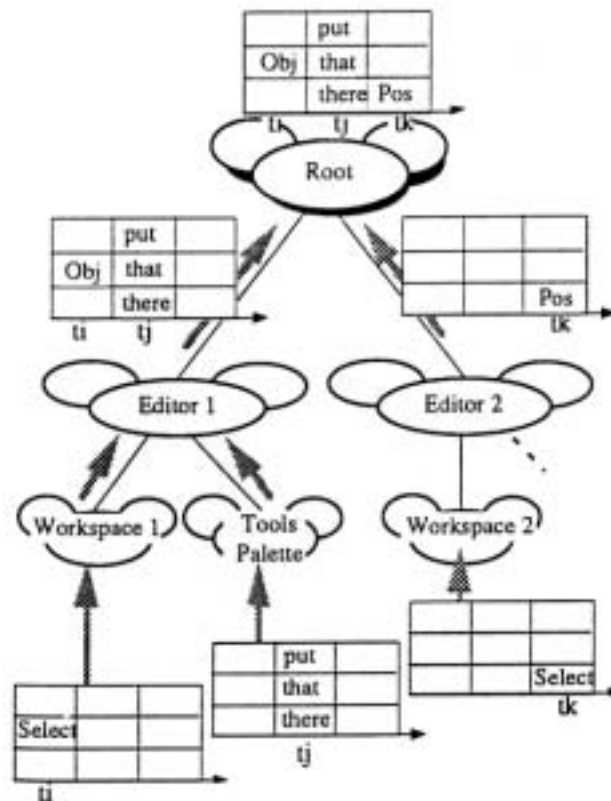


Figure 5. Composing melting pots in a system similar to Bolt's [1980] Put-That-There with PAC-Amodeus [Nigay and Coutaz, 1993].

The system presented in Figure 5 uses a two-level fusion process for a graphics editor that supports speech and mouse gesture. In this example, the user says, "put that there" and at the same time uses the mouse to select the object to be moved and to indicate the destination in a distinct workspace. The

named ellipses in the figure represent agents. Workspace agents interpret the events that occur in the drawing areas. The tool palette agent is responsible for the events issued in the palette. Editor agents combine information from the lower levels into higher abstractions. Finally, the Root agent performs a second level of fusion to obtain a complete command, and sends it to the functional core.

In their 1995 paper Nigay and Coutaz explain the fusion mechanism in detail and give the metrics and rules that they use in the fusion process. They divide fusion in three classes. *Microtemporal fusion* is used to combine related informational units produced in a parallel or pseudo-parallel manner. It is performed when the structural parts of the input melting pots are complementary and when their time intervals overlap. *Macrotemporal fusion* is used to combine related information units that are produced or processed sequentially. It is performed when the structural parts of the input melting pots are complementary and their time intervals belong to the same temporal window but do not overlap. *Contextual fusion* is used to combine related information units without attention to temporal constraints.

Their algorithm favors parallelism and is thus based on an eager strategy: it continuously attempts to combine input data. This may possibly lead to incorrect fusion and requires an undo mechanism to be implemented. This architecture has been used to build MATIS [Nigay *et al.*, 1993], a multimodal airline travel information system that uses speech, direct manipulation, keyboard and mouse or combinations of them as its input.

Open Agent Architecture

Moran *et al.* [1997] present another agent-based model for synchronizing input modes in multimodal interaction. They have implemented multimodal interfaces as an application of the Open Agent Architecture (OAA). The OAA is a general agent platform that supports distributed and mobile agents that can react on multiple input modalities. A *modality coordination agent* (MC agent) is responsible for combining the input in the different modalities to produce a single meaning that matches the user's intention.

The MC agent uses a simpler fusion mechanism than Nigay and Coutaz [1995]. If information is missing from the given input, the MC agent searches the conversation context for an appropriate reference. If it can not be found, the agent waits for a certain amount of time (2-3 seconds) before asking the user to give additional information. This time delay is designed to handle the

synchronization of different input modalities that require a different amount of time to process the input.

The OAA has been used to implement a wide range of systems, for example, an office assistant that helps the user in information retrieval, a map-based tourist information system that uses speech and gesture, and an air travel information system that is used through a Web and telephone interface.

Modeling the fusion of multiple modalities is an important problem in multimodal interaction. Both PAC-Amodeus and the Open Agent Architecture have been successfully used to implement prototype systems, and these architectures can support a wide range of modalities. However, the fusion of different input modalities offers plenty of research opportunities.

2.3. The history of multimodal user interfaces

Many types of multimodal user interface prototypes have been built to demonstrate the usefulness of multimodal interaction. In this section I present a summary of these systems.

Virtual reality systems have been studied since Morton Heilig's Sensorama [Heilig, 1955]. These systems are similar to multimodal user interfaces in their use of many parallel modalities for both input (i.e., datagloves, speech, and exoskeletons) and output (i.e., head-mounted displays, three-dimensional audio, and tactile feedback). However, virtual reality systems are also quite different from multimodal user interfaces, and can thus be viewed as a separate line of research that is useful for the researchers and developers of multimodal user interfaces, because both fields need better input and output devices. The real difference between these systems was nicely explained in [MIAMI, 1995, p. 7]: "Virtual reality aims at the imitation of reality by establishing immersive audio-visual illusions, whereas multimodality attempts to enhance the throughput and the naturalism of human-computer interaction."

Bolt [1980] introduced the notion of the multimodal user interface in his Put-That-There system (Figure 6). In this system the user could move objects on screen by pointing and speaking. The user pointed at the object, said, "put that", pointed at the destination and said "there." This system combined pointing and speaking in a single action. Later, more sophisticated prototypes have been developed by adding eye tracking in a similar system [Bolt and Herranz, 1992; Thorisson *et al.*, 1992; Koons *et al.*, 1993]. This additional input modality can help to understand the user if the pointing gesture is not clear. In normal human-human communication, humans express themselves in many

simultaneous and redundant ways [Feldman and Rimé, 1991]. Several parallel input modalities help building user interfaces less moded than interfaces using only one input modality.



Figure 6. Bolt's Put-That-There system [Bolt, 1980].

Cohen *et al.* [1989] presented two systems, CHORIS and SHOPTALK that make use of combining direct manipulation with natural language. In their systems natural language is entered using a keyboard. They argued that the use of these modalities overcomes some of the limitations of each. The synergy of these modalities comes from joining the predictability and clarity of direct manipulation with the descriptive power of natural language. Wahlster [1991] presented a similar system, XTRA, as a multimodal interface to expert systems.

CUBRICON [Neal *et al.*, 1989; Neal and Shapiro, 1991] is a system that uses mouse pointing with speech. Cohen *et al.* [1997] have implemented a prototype called QuickSet that integrates speech with pen input that includes drawn graphics, symbols, gestures and pointing.

Weimer and Ganapathy [1989] described a multimodal environment that uses gestures and speech input to control a CAD system. They used a dataglove to track the gestures and polarizing glasses to present the objects in three dimensions.

Oviatt [1996] presented a multimodal system for dynamic interactive maps. In her system speech and pen writing are used either separately or multimodally to perform map-based tasks. Oviatt *et al.* [1997] found out that spoken and written modes in the system consistently supplied complementary semantic information rather than redundant information.

Yang *et al.* [1998] present a group of visual tracking techniques that can be used in multimodal user interfaces. They have implemented a camera-based face tracker that can also track face parts like eyes, lips and nostrils. The system can also estimate the user's gaze direction or head pose. They have implemented two multimodal applications that make use of these techniques: a lip-reading system and a panoramic image viewer. The lip-reading system improves speech recognition accuracy by using visual input to disambiguate among acoustically confusing speech elements. The panoramic image viewer uses gaze to control panning and tilting, and speech to control zooming.

Public kiosks are promising applications for multimodal interfaces. Even though most current public information kiosks are limited in their interaction with the user, some experimental multimodal information kiosk prototypes have been developed. Probably the most advanced kiosk prototype is the Digital Smart Kiosk project [Christian and Avery, 1998]. This kiosk uses computer vision and touchscreen input and presents the information on graphics display that is augmented with a talking head, DECface [Waters and Levergood, 1993].

2.4. Potential benefits of multimodal interfaces

Maybury and Wahlster [1998, p. 15] give a list of potential benefits of multimodal user interfaces: efficiency, redundancy, perceptability, naturalness, accuracy, and synergy. *Efficiency* follows from using each modality for the task that it is best suited for. *Redundancy* increases the likelihood that communication proceeds smoothly because there are many simultaneous references to the same issue. *Perceptability* increases when the tasks are facilitated in spatial context. *Naturalness* follows from the free choice of modalities and may result in a human-computer communication that is close to human-human communication. *Accuracy* increases when another modality can indicate an object more accurately than the main modality. Finally, *synergy* occurs when one channel of communication can help refine imprecision, modify the meaning, or resolve ambiguities in another channel.

In the following section I will discuss three important benefits of multimodal user interfaces: natural dialogue, efficient dialogue, and granularity. The first two benefits are discussed using a public information kiosk as an example. Natural and efficient dialogue is possible because the users have many options to carry out different tasks, and granularity of the system follows from mutual disambiguation of recognition errors. Another example of a multimodal system is discussed in the next section with some common misconceptions about multimodal user interfaces.

2.4.1. Natural and efficient dialogue

Public information kiosks have a wide range of different users that may or may not have basic computing skills. Thus, these systems should be usable for all kind of users. In addition, the kiosks are typically used within a very short time, and therefore they should not require extensive training or long manuals.

Many simple kiosk systems have physical buttons and small keyboards as the only controls. These controls seem to work well in simple and constrained uses. Touchscreens alone can perform well in many kiosks. If the system is designed to be used with a button-based interface as the Digital's Smart Kiosk prototype [Christian and Avery, 1998], it can be effectively controlled with a touchscreen.

The real power of Digital's Smart Kiosk prototype is in its multimodal use of machine vision and touchscreen input. Machine vision is able to determine when a user is approaching the kiosk, and also whether he or she stops in front of the kiosk or walks by it. If the user slows down or stops in front of the kiosk, the DECface talking head [Waters and Levergood, 1993; Parke and Waters, 1996] welcomes the user and the kiosk presents different options as touchable buttons on screen. Machine vision is further used to track the user and control the talking head so that it will look at the user.

In general, different modalities can be used to supply either redundant or complementary information. Redundant information helps in determining the goal of the user if it is not clear from just one modality. Complementary information enables more efficient and natural interaction. A multimodal user interface is – at its best – a natural and convenient way to interact with computing systems. When a diverse group of users is expected to use the kiosk system and no training can be given beforehand, many different modalities can be used to track the user and model what his or her intentions are. Many redundant ways to carry out operations make the use more natural since the user can choose the most natural input method for a given task.

A multimodal user interface is efficient when the user chooses an efficient input modality for the specific task. This benefit also follows from redundancy, since there are several ways to give the same input. Complementary modalities make the system even more efficient when two parallel or sequential input modalities make the task easier and more efficient than using a single modality, such as speech. Oviatt *et al.* [1997] have observed this advantage in empirical studies.

2.4.2. Mutual disambiguation of recognition errors

Many recognition techniques produce frequent recognition errors. This is the case with speech recognition, eye tracking, head tracking, lip-reading and many other machine vision techniques. In fact, the only truly accurate techniques are usually based on mechanic structures, such as so-called exoskeletons [Pimentel and Teixeira, 1995] in virtual reality systems. When we attach devices onto users, they can be monitored quite accurately. But the goal in building natural multimodal user interfaces is to build the interfaces as seamless as possible so that nothing needs to be attached to the users. Thus recognition errors are unavoidable.

Promising results on using audio-visual information to improve speech recognition accuracy have been presented in psychology [McGurk and MacDonald, 1976; MacDonald and McGurk, 1978; Sams *et al.*, 1998] and in some experimental computing systems [Yang *et al.*, 1998]. Many lip-reading systems require the users to keep still or put special marks on their faces. The research on these systems has also progressed. Yang *et al.* [1998] have recently presented a computer vision based lip-reading system that does not require the presence of such hindrances.

Audio-visual integration is not the only multimodal combination that can help increase the granularity of a system. Oviatt [1999a] uses her multimodal interface for dynamic interactive maps [Oviatt, 1996] to determine how multimodal gestures can help in speech recognition with native and accented speakers of English. She found that although stand-alone speech recognition performed far more poorly for accented speakers, their multimodal recognition rates did not differ from those of native speakers. This gives strong evidence for the benefits of combining many error-prone recognition technologies in a multimodal system that may in result have better granularity than these technologies have when used separately.

2.5. Common misconceptions on multimodal interfaces

The problems with multimodal interfaces may arise from false expectations. Often the benefits of multimodality are taken for granted, and the importance of delicate design is not taken into account. Sharon Oviatt [1999b] has explored ten common “myths” of multimodal interaction. She declares these myths as misconceptions and presents contrary empirical evidence for them. The evidence comes mostly from empirical testing of Oviatt’s [1996] multimodal interactive maps that make use of speech recognition and pen input with drawn graphics, symbols, gestures, and pointing.

Next I will summarize Oviatt’s empirical evidence that justifies calling these claims as myths. The ten myths were the following:

1. *If you build a multimodal system, users will interact multimodally.* Even though the users prefer to interact multimodally, at least in spatial application domains [Oviatt, 1997], there is no guarantee that they will issue every command to a system multimodally. They typically intermix unimodal and multimodal expressions. Specifically, whether a user will express a command multimodally depends on the type of action he or she is performing. The users almost always expressed the commands multimodally when describing spatial information about the location, number, size, orientation, or shape of an object [Oviatt *et al.*, 1997], and often used multimodal commands when selecting an object from a larger array, but almost not at all when they performed general actions without any spatial component.
2. *Speech and pointing is the dominant multimodal integration pattern.* Oviatt *et al.* [1997] found that a speak-and-point pattern only comprised 14% of all spontaneous multimodal expressions. Pen input replaced speech in many tasks. The users are likely to choose the modality that best suits for a given task.
3. *Multimodal input involves simultaneous signals.* Oviatt *et al.* [1997] also found that about half the time the pen input and speech were sequentially integrated, and that gesturing often preceded speech with a brief lag of one or two seconds between input signals. To summarize, even though speech and gesture are highly interdependent and synchronized during multimodal interaction, synchrony does not imply simultaneity.
4. *Speech is the primary input mode in any multimodal system that includes it.* Oviatt [1997] noticed that other modes could convey information that is

not present in speech signal at all, for example, spatial information specified by pen input. Thus, speech should not be considered as the exclusive carrier of important content. The problem is that thinking speech as the primary channel risks underexploiting the other modalities.

5. *Multimodal language does not differ linguistically from unimodal language.* Oviatt [1997] found that multimodal pen/voice language is briefer, syntactically simpler, and less disfluent than users' unimodal speech. When free to interact multimodally, users selectively eliminate many linguistic complexities. For example, they prefer not to speak error-prone spatial location descriptions if a more compact and accurate alternative is available.
6. *Multimodal integration involves redundancy of content between modes.* Oviatt [1999b] claims that the dominant theme in users' natural organization of multimodal input is complementary of content, not redundancy. This was noticed in [Oviatt *et al.*, 1997].
7. *Individual error-prone recognition technologies combine multimodally to produce even greater unreliability.* This claim was already discussed in the previous section. The increased robustness is due to leveraging from users' natural intelligence about when and how to deploy input modes effectively. People will avoid using an input mode that they believe is error-prone for a certain action. When a recognition error does occur, users alternate input modes to better be able to complete the task. The increased robustness of unreliable technologies requires that the system integrates the modes synergistically, as in Oviatt [1999a].
8. *All users' multimodal commands are integrated in a uniform way.* Different users can have different integration patterns [Oviatt 1999b]. This emphasizes the need for adapting a multimodal user interface to a group of users. If a set of different integration patterns can be found, the system could adapt to a user when it recognizes this user's dominant integration pattern.
9. *Different input modes are capable of transmitting comparable content.* It is delusive to think that all input modalities are interchangeable. They all have differences in the type of information they transmit, their functionality during communication, the way they are integrated with other modes, and their basic suitability to be incorporated into different interface styles. Most likely, a simple one-to-one translation is not possible without losing something: content, accuracy, or ease of use.

10. *Enhanced efficiency is the main advantage of multimodal systems.* A modest speed enhancement by 10% was documented in [Oviatt, 1997] in a spatial domain. However, the same study showed that task-critical errors and disfluent language could drop by 36% – 50% during multimodal interaction. In addition, we should not underestimate the flexibility that results from permitting the users to alternate between different input modes. This flexibility makes it easier to develop user interfaces that are suited for all kinds of users in different environments.

The reason for presenting these common misconceptions in such length is to emphasize the need for careful design of multimodal user interfaces. Oviatt's [1999b] article gives many relevant issues to consider when multimodal user interface architectures are designed. The success of a multimodal system may depend on a very small detail: a detail that the people that implement the system consider as self evident.

2.6. Guidelines for multimodal interaction

Multimodal user interfaces form a very large group of different user interfaces. As described above, these systems may make use of a wide variety of different input and output modalities, and mix them freely. Appreciating the vast potential that multimodality has in making the human-computer interaction more natural, easier and even more efficient, it becomes even more important to be able to define guidelines for *where and when to use which modalities*. Unfortunately, this is a vast research issue that can not be answered within this dissertation. Thus, I have selected a different approach, which argues for the research issues that should be addressed by researchers of human-computer interaction and psychology, side by side, in order to gain better understanding of applying multimodal interaction in computing systems.

A motivating example can be found in [MIAMI, 1995, p. 43]. They present a question: what happens if more human output channels are used in human-computer interaction? They had two hypotheses:

Hypothesis (1): The combination of human output channels effectively increases the bandwidth of the human→machine channel. This has been discovered in many empirical studies of human-computer interaction [Oviatt, 1999b].

Hypothesis (2): Adding extra output modality requires more neurocomputational resources and will lead to deteriorated output quality, resulting in reduced effective bandwidth. Two types of effects

are usually observed: a slow-down of all output processes and interference errors due to the fact that selective attention cannot be divided between the increased number of output channels. Examples of this are writing when speaking and speaking when driving a car.

So the main question is: do we know how to build a user interface for which hypothesis 1 is true and hypothesis 2 is not. The current answer is: we do not. There is some encouraging empirical evidence [Oviatt, 1997; Oviatt *et al.*, 1997; Oviatt, 1999a; Oviatt, 1999b] that spatial information can be entered with pen gestures resulting in great efficiency when used with speech. Other similar studies exist and report on a specific combination of input modalities that was used in a prototype system. However, there is no general understanding of which modalities to use for which task.

Cognitively, it is possible to find out which tasks require more cognitive resources and thus are more difficult to do in parallel with the other tasks. The understanding of the brain is important in order to find out the relationships between human input and output modalities. We know great many things of the brain but they are not understood well enough for us to be able to assign the right modalities to given tasks.

Still, we should assign specific computer input devices to carry out these tasks. New devices are developed all the time and there already is a very large collection of different devices with varying accuracy, ergonomics and demand on processing power. Even when the right devices have been found we need to decide how to use them effectively in a given system.

To summarize, more research is needed to understand the following:

- *how the brain works and which modalities can best be used to gain the synergy advantages that are possible with multimodal interaction,*
- *when a multimodal system is preferred to a unimodal system,*
- *which modalities make up the best combination for a given interaction task,*
- *which interaction devices to assign to these modalities in a given computing system, and*
- *how to use these interaction devices, that is, which interaction techniques to select or develop for a given task.*

Interdisciplinary cooperation is essential in order to get better understanding on multimodal interaction. A result that two modalities are intertwined and that there are multimodal receptor cells in the brain does not yet give any advice on *how and when* to use this interesting result in human-computer interfaces.

A common notion in cognitive psychology is that there are several codes in which information is processed, as visual-spatial and verbal codes [Baddeley, 1986]. The codes allow us to define the manner by which the information will be processed. A multiple resource theory can be constructed with the assumption that each type of information code has its own processing capacity and its own processing resources [Allport, 1980; Wickens, 1992]. A model of human information processing with several stages and codes can then be proposed [Wickens, 1992, p. 375].

Baber [1997, pp. 271-275] discusses multiple resource theory research that would answer many of the questions that I presented above. He points out many problems with the current research on this theory:

- (1) the mechanism by which tasks compete or combine is difficult to define,
- (2) using secondary tasks together with primary tasks complicates the problems of deciding whether the results arise due to competition, compatibility or scheduling of attentional demands across the tasks,
- (3) definition of codes for certain modalities may be difficult, and
- (4) currently multiple resource theory only considers visual and auditory modalities for perception; it is not clear whether this theory can be applied to other modalities.

After discussing the problems that exist in understanding multimodal interaction, one can see the enormous amount of research that still needs to be done in order to provide comprehensive guidelines for multimodal human-computer interaction. This will certainly require a long time, a great number of people and great investments in research.

A modality theory has been developed in a large European research project [AMODEUS, 1995]. The theory is comprehensive for choosing and combining output modalities, but the project had only begun a similar work with input modalities when its seven-year financing period ended. However, the fact that they recognized the need for a modality theory for input modalities suggests that there is wider interest in developing such a theory.

2.7. Summary

Different aspects of multimodal interaction were discussed in this chapter. There still are different views on how to define multimodal user interfaces. In general, multimodal user interfaces may involve combined outputs, as speech and animation in a talking head, or combined inputs as in Bolt's [1980] Put-That-There system that was based on speech recognition and hand gestures.

Multimodal interaction offers many potential benefits: natural dialogue, efficient interaction, and disambiguation of recognition errors. In order to get these benefits the systems need to be designed well and the modalities need to be either redundant or complementary. Redundant modalities allow the user several ways to carry out an operation. In addition, they allow the system to determine the intention of the user if one or more of the inputs would not be unambiguous separately. Complementary modalities each add some synergistic value when the user is carrying out a single task, therefore making the operation more efficient and also more natural, when the system is designed well.

I believe that multimodal interaction will be a prominent interaction paradigm that may eventually replace the current graphical user interface paradigm that is based on the use of windows, icons, menus and a pointing device (WIMP). Many authors have suggested this with different names: Non-command user interfaces [Nielsen, 1993], Post-WIMP user interfaces [van Dam, 1997] and Non-WIMP user interfaces [Jacob *et al.*, 1999].

In this dissertation I have adopted the system-centered view by which a multimodal user interface increases the user's degrees of freedom by allowing many simultaneous inputs. I present two prototype systems and a group of multimodal interaction techniques that make use of two pointing devices, speech and touch.

3. Two-handed interaction

Graphical user interfaces have become the industry standard in current software. Even though they were introduced in Xerox Star workstations [Lipkie *et al.*, 1982; Bewley *et al.*, 1983; Baecker and Buxton, 1987] almost two decades ago, graphical user interfaces have developed surprisingly little after that. Small improvements have appeared in commercial operating systems, such as better feedback and more direct ways to perform frequent tasks. In this chapter I discuss two-handed interaction, an improvement to the ways we use in controlling these interfaces. Two-handed interaction can also be described as a special case of multimodal interaction that makes use of both hands operating in parallel.

3.1. The psychology of two-handed interaction

Human bimanual control has been extensively analyzed in psychology. Much of this research specializes in defining which parts of the brain control which hand, and how to determine the handedness of the subjects. Many of these results are not directly applicable to building user interfaces, but there are some useful theories that explain the differences between the hands and the way the two hands cooperate in bimanual tasks.

3.1.1. Analyzing manual tasks

Fitts [1954] argued that the speed with which manual movements are made is limited by the information-processing capacity of the motor system. He presented a law that explains how the difficulty of a tapping task varies based on movement time and amplitude. He based his model on an analogy to Shannon's theorem 17 [Shannon and Weaver, 1949, pp. 100-103]. The rate of information processing (index of performance, *IP*) of the hand-arm is analogous to channel capacity in Shannon's theorem and can be calculated according to the formula

$$IP = ID / MT,$$

where ID is the index of difficulty (in bits) and MT is the movement time (in seconds).

Fitts suggested that the width of the target (W) is analogous to Shannon's noise and the amplitude of the movement (A) is analogous to Shannon's electronic signals, and that they together specified the index of difficulty (ID) of the movement, according to the formula

$$ID = \log_2 (2A / W) \text{ bits.}$$

The index of difficulty is measured in "bits" and describes the difficulty to carry out the task. The greater the ID value is the more difficult the task is.

A more general way to present Fitt's law is the following regression equation:

$$T = a + b \log_2 (2A / W)$$

where a and b are empirically determined regression coefficients. The reciprocal of b is the index of performance (IP) that is measured in bits/second. Detailed explanation and analysis of Fitts' law, its variations and its use in human-computer interaction can be found in [MacKenzie, 1992]. One remarkable result of using Fitts' law in human-computer interaction is that it makes it possible to compare input device performance results from different studies using the index of performance. However, MacKenzie [1992] found that these comparisons are still difficult because of task differences, selection techniques, range of conditions employed, and dealing with response variability.

Fitts' law predicts a tradeoff in human movement: the faster we move our hands, the less precise our movements are. The law applies in one kind of tasks: hitting a target over a certain distance. In human-computer interaction it is suited for analyzing pointing and dragging tasks.

3.1.2. Differences between the preferred and non-preferred hand

There are three theories that explain the differences of the two hands in rapid aimed movements [Annett *et al.*, 1979; Kabbash *et al.*, 1993].

The first is that preferred and non-preferred hands differ primarily in their use of sensory feedback. Flowers [1975] defined so-called ballistic and controlled movements. Ballistic movements last less than 400 ms and they are too quick to make use of sensory feedback. Controlled movements benefit from visual and tactile sensory feedback. Flowers found that the preferred and non-preferred hands of strongly lateralized subjects achieved equal rates in a rhythmic tapping task, but that in a variation of Fitts' reciprocal tapping task

(ID ranging from 1 to 6) the preferred hand outperformed the non-preferred hand by 1.5 – 2.5 bits/s, with differences marked at all but the lowest two IDs.

A second theory [Schmidt *et al.*, 1979] explains that the preferred hand is less noisy in its output function. According to this theory, increases in amplitude or decreases in movement time require a greater force, which leads to greater output variability and more errors. The theory suggests that variability in force amplitude varies inversely with the square of movement time and the variability of force duration varies directly with movement time. Annett *et al.* [1979] noticed that the non-preferred hand makes nearly 50% more corrective movements than the preferred hand, and that these movements last on average about 120 ms.

A third theory by Todor and Doane [1978] predicts a non-preferred hand advantage for larger target distances. The theory is based on Welford's [1968] proposal that rapid aimed movements are composed of two distinct parts: a fast distance-covering phase and a slower target homing phase. The first phase is similar in speed to ballistic movement, but the second phase requires visual control. They hypothesized that within tasks of equivalent calculated difficulty, movement time with the preferred hand should increase as the width and amplitude of the target grows larger while it should decrease with the non-preferred hand. Obviously both time changes are in limited scale. Kabbash *et al.* [1993] carried out a study that supports this theory and extends Todor and Doane's results to a wide range of task difficulties. Specifically, in tasks of equivalent difficulty, between-hand comparisons showed a right-hand advantage in movement time for target width, but a left-hand advantage for amplitude.

None of these theories makes explicit predictions about differences between preferred and non-preferred hands. However, even in their current form they are useful when deciding what each hand should do in the user interface. According to Flowers [1975] the non-preferred hand can be used for simpler tasks than the preferred hand. These tasks should not require much sensory feedback. Schmidt *et al.* [1979] emphasize the inaccuracy and error-proneness of the non-preferred hand. Thus the non-preferred hand should not be used in tasks that require high accuracy. According to Todor and Doane [1978], there may be tasks in which the non-preferred hand actually performs better than the preferred hand. These tasks should contain long distances and large objects.

3.1.3. Division of labor between the hands

To be able to better understand bimanual interaction, Guiard [1987] has presented a kinematic chain theory. According to his model, the functions of both hands are related serially so that the non-preferred hand acts as a base link and the preferred hand as the terminal link. This theory suggests cooperative and asymmetric function of the two hands. The basic characteristics of the kinematic chain model are the following (assuming a right-handed person):

1. Right-to-left spatial reference: the right hand typically finds its spatial references in the results of the motion of the left hand. Often the non-preferred hand plays a postural role in keeping an object steady while the preferred hand executes a manipulative action on it. For example, when the left hand holds a nail and the right hand uses a hammer.
2. Left-right contrast in the spatial-temporal scale of motion: the movements of the left hand usually have a low temporal frequency (long periods between initiating the movements) and a low spatial accuracy (large movement amplitudes) compared to the high-frequency and detailed work of the right hand. The preferred hand is capable of producing finer movements than the non-preferred. For example, an artist holds the palette in her left hand and paints with the right hand.
3. Left-hand precedence in action: usually the action starts with the non-preferred hand and ends with the preferred hand. For example, a sheet of paper is grabbed with the left hand and the right hand is used to write on it.

The previous examples are natural two-handed tasks. However, drawing with the left hand while writing with the right hand is not natural at all. We can assume that natural two-handed tasks belong to synergistic multimodal interactions [Nigay and Coutaz, 1993] and have smaller cognitive demands than other two-handed tasks.

3.1.4. Cognitive advantages in two-handed interfaces

Leganchuk *et al.* [1998] found that in addition to motor advantages, two-handed interfaces also have cognitive advantages over one-handed interfaces. They refer to *chunking* [Simon, 1974] which means that the acquisition of skills is based upon the compilation and proceduralization of knowledge into chunks formed around understanding subtasks. Two-handed input may allow the users to perform at a natural level of chunking that corresponds to daily activities, possibly resulting in the following interrelated cognitive advantages:

1. Reduce and externalize the load of planning/visualization in unimanual input. Because two-handed input allows the user to treat the task as a larger and more natural gestalt, the user no longer needs to compose, think, and plan the elementary steps of a task.
2. Rapid feedback of manipulation results in a higher level of task: the user immediately sees the result of action in relation to the goal state.
3. Support epistemic action: the user may take advantage of the two-handed input and perform actions of an epistemic nature in addition to those of a pragmatic nature.

They have made the above observations in a number of two-handed tasks. Leganchuk *et al.* [1998] also found that the benefits of two-handed interfaces increase as it becomes more difficult to mentally visualize the task, which supported their hypothesis that there is also a cognitive advantage in two-handed interaction over one-handed interaction.

However, other studies [Kabbash *et al.*, 1994; Balakrishnan and Kurtenbach, 1999] have shown that increased cognitive load can result in reduced performance in some bimanual tasks. These results emphasize the requirement that two-handed interaction needs to be carefully designed.

3.2. Modeling two-handed interaction

The need for designing two-handed interfaces has driven the development of new modeling techniques that take into account many simultaneous inputs. Myers [1990] presented a model for handling input that has been implemented in the Garnet user interface management system. His model encapsulated interactive behaviors into a few “Interactor” object types that hide the underlying window manager events. A Garnet interactor can support multiple input devices operating in parallel. They are similar to the controllers in the Smalltalk “model-view-controller” (MVC) paradigm [Krasner and Pope, 1988].

Chatty [1994] extended a graphical toolkit for two-handed interaction. He discusses the extensibility of the input system of graphical toolkits and proposes some abstractions for describing combined interactions. His Whizz toolkit uses a data-flow model. All the objects manipulated in Whizz are modules that can be connected together with links that carry pieces of information. A fusion of events is represented much in the same way as in Myers’ Garnet with graphical objects that have two inputs and a fused output.

Buxton [1990] presented a three-state model of graphical input. In Buxton’s model the devices can be in three states: out of range (0), tracking (1), and

dragging (2). Figure 7 shows a three-state model for the puck on the Wacom tablet. A puck is a special device that has its own identifier and can be detected on top of the tablet.

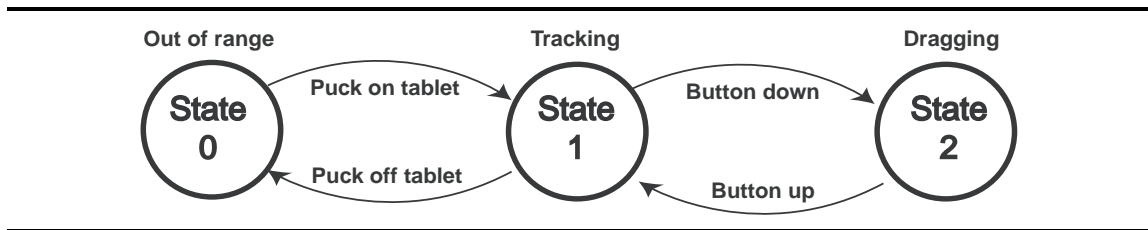


Figure 7. State transitions for the puck on the Wacom tablet [Hinckley *et al.*, 1998b].

Figure 8 shows the transitions for a touchpad and a standard mouse. The touchpad only senses states 0 and 1, while a standard mouse only senses states 1 and 2. However, the touch-sensing mouse [Hinckley *et al.*, 1998b] and trackball [Hinckley and Sinclair, 1999] have all three modes since they recognize when the user touches the device. Touchpads can also have all three modes if they include integrated buttons.

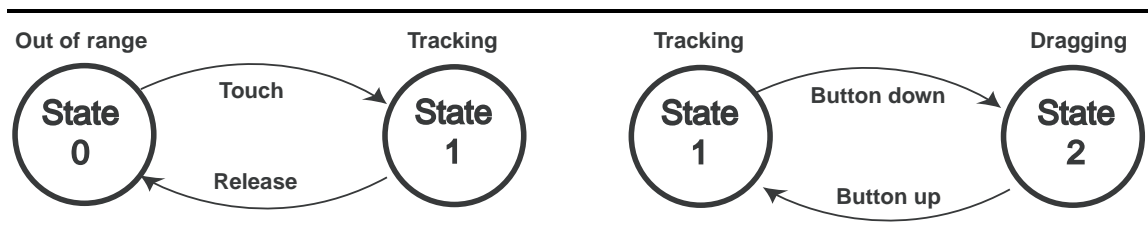


Figure 8. State transitions for a touchpad (left) and a standard mouse (right) [Hinckley *et al.*, 1998b].

Hinckley *et al.* [1998b] proposed several enhancements to Buxton's model. First, they added notation for the continuous properties of the input devices. The notation is essentially the same that Mackinlay *et al.* [1991] described. Figure 9 presents this notation.

Notation	Description of property sensed
x	absolute position sensing
dx	relative position sensing (motor sensing)
R	absolute angular
dR	relative angular
F, dF	force and change in force
T, dT	torque and change in torque
nil	state cannot sense any continuous input

Figure 9. Notation for continuous properties by Hinckley *et al.* [1998b].

Second, Hinckley *et al.* [1998b] distinguished between out-of-range states based on touch versus those based on device proximity to be able to model their new touch-sensitive devices. To accomplish this they propose that state 0 is actually a union of two potentially different states, T0 (out-of-range state triggered by touch) and P0 (out-of-range state triggered by device proximity to a sensor). To be able to model two-handed interaction techniques they doubled the states and added subscripts p (preferred hand) and n (non-preferred hand) in each of the states $P0_p$, 1_p , 2_p , $P0_n$, 1_n and 2_n . This solution was not sufficient because the model becomes too complicated [see Hinckley *et al.*, 1998b, Fig. 10] and it does not express the inherent parallelism of two hands. This is why they presented a new model that is a special case of a Petri net model.

The new Petri net model preserves much of the flavor of the three-state model but also enables explicit representation of the parallelism. The model is used to represent composite two-handed interaction techniques, not specific capabilities of the individual devices. Figure 10 shows an example on a simple Petri net model for a tablet with a puck and a stylus that are used in separate tasks. There are two circular tokens, a black token representing the preferred-hand device, and a gray token representing the device in the non-preferred hand. The black token can only visit states with a solid border and the gray token can only visit states with a dashed border. A token can traverse any arc as long as it leads to a state that the token is allowed to visit. The coloring of the arcs (solid or dashed) indicates which input device generates the signal corresponding to that arc.

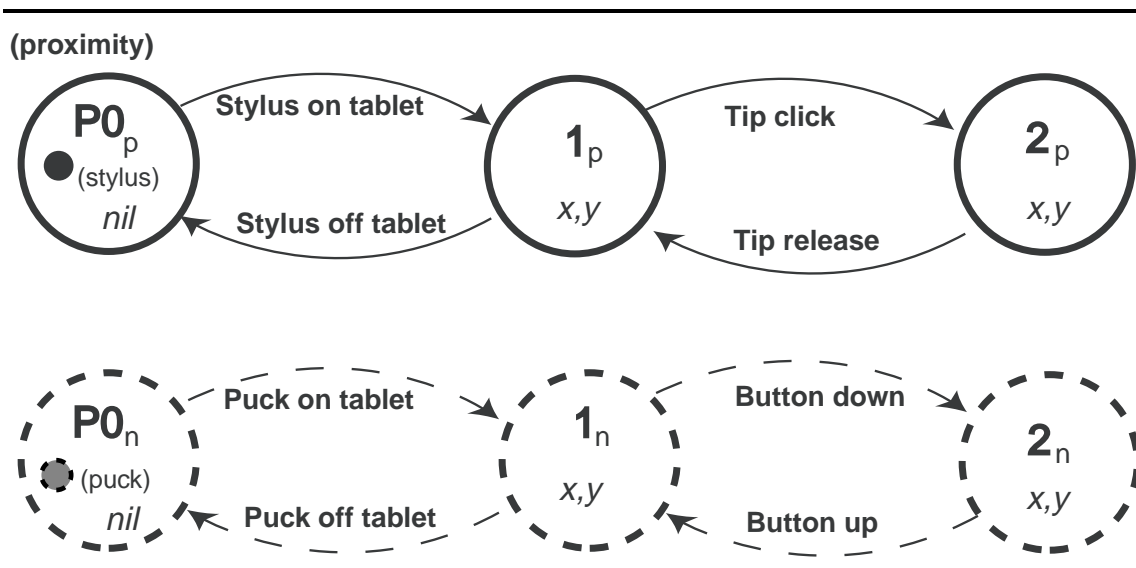


Figure 10. A simple Petri net model for puck and stylus on the tablet [Hinckley *et al.*, 1998b].

Figure 11 is a refinement of Figure 10 and adds a two-handed dragging technique that requires using both input devices. There is an additional state 2_{np} , which is the only state in the diagram that both the black and gray token can enter. That is why it is colored with both a solid and dashed border. If the system is in state 2_{np} and the mouse button is released, the gray token returns to state 1_n . Similarly, the black token returns to state 1_p . The tokens change to a new state separately as the corresponding input devices are used. Still, it can be seen in Figure 11 that if one of the tokens moves out of the combined state 2_{np} , the other returns to its normal state 2 following the corresponding arc.

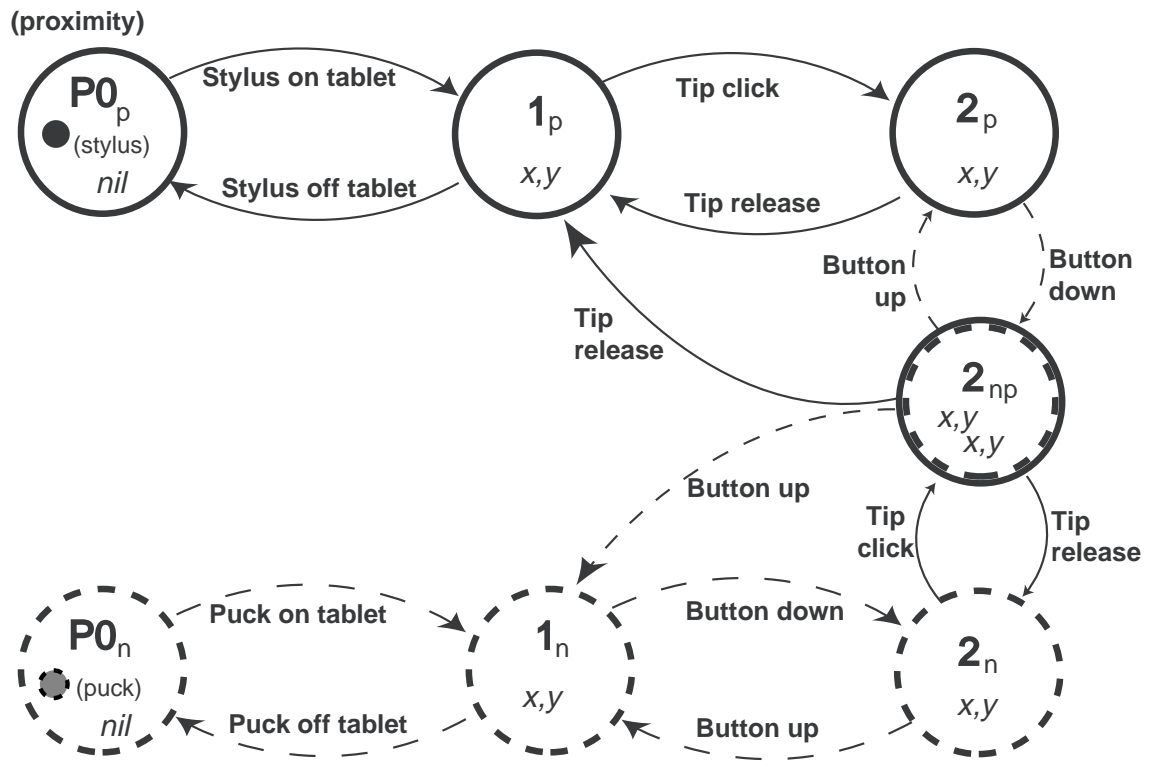


Figure 11. A complete Petri net model for puck and stylus on the tablet [Hinckley *et al.*, 1998b].

Jacob *et al.* [1999] have developed a programming-oriented software model for Non-WIMP user interfaces. They describe “Non-WIMP” interfaces as having parallel continuous input with the user, and two-handed interaction can also be modeled with this more general model. The model combines a data-flow or constraint-like component for the continuous relationships with an event-based component for discrete interactions, which can disable or enable individual continuous relationships.

3.3. The history of two-handed user interfaces

The non-preferred hand has been used as early as in Ivan Sutherland's [1963] Sketchpad system in a discrete task to press buttons that had an effect on what a light pen on the preferred hand would do when the user drew on the computer screen with it. Practically all graphical user interfaces use the non-preferred hand in a similar way when the user presses a modifier key (control, alt, shift) at the same time as he or she selects objects with the mouse. Leganchuk *et al.* [1998] mention video games that often have controllers that require pressing buttons at the same time as the position is controlled with some kind of joystick. This kind of control requires skill and good coordination of the two hands, but the users accept it since an attractive video game motivates them to do it and two-handed control helps them in playing.

Qwerty-typing is another exception of general one-handed interactions, but it is not like our everyday bimanual actions: each finger performs a discrete task of pressing a button. However, keyboard shortcuts help an experienced user speed up the use of graphical user interfaces by allowing the non-preferred hand to replace a menu selection command with a key press.

To my knowledge, Krueger [1983, pp. 142-147; Krueger *et al.*, 1985] was the first to suggest two-handed computing applications that use both hands to give continuous input in a coordinated and integrated manner. He proposed several two-handed gestures that could be used in text editing and two- and three-dimensional graphics. Matsushita and Rekimoto [1997] have implemented the same kind of interaction techniques as Krueger [1983] suggested in their HoloWall prototype.

The first evaluation of two-handed interfaces was done by Buxton and Myers [1986]. They implemented two-handed desktop systems that were used in graphical manipulation and word processing. The second experiment of Buxton and Myers [1986] was repeated by Zhai *et al.* [1997] with many different input device configurations.

Kabbash *et al.* [1994] showed that two-handed interfaces need to be designed well in order to get benefits from them. In their study, four different interaction techniques were compared. The fastest technique was a two-handed ToolGlass-technique that was based on earlier work on toolglasses and magic lenses [Bier *et al.*, 1993]. However, the slowest technique was a two-handed technique that used two separate cursors. The normal one-handed technique performed better than the two-handed technique even though it had the same manual benefits – shorter distances and division of labor between the hands –

as in the second task of Buxton and Myers [1986]. The task that Kabbash *et al.* used was similar to the task that Dillon *et al.* [1990] used in a similar comparison. There, the two-handed setup did not give significant benefits over one-handed techniques.

Fitzmaurice and Buxton [1997] and Fitzmaurice *et al.* [1995] have implemented two-handed techniques in graspable user interfaces. Graspable user interfaces are tangible and real since the controls are real bricks that the user manipulates over a tablet. This line of research has continued in MIT's metaDESK system [Ishii and Ullmer, 1997; Ullmer and Ishii, 1997].

Toolglasses and magic lenses [Bier *et al.*, 1993; Bier *et al.*, 1994; Kabbash *et al.*, 1994] is a desktop metaphor in which the tools are controlled with both hands. Tools are drawn on a semitransparent palette that is normally moved by rolling a trackball with the non-preferred hand. A tool or action is selected when the user clicks through it in an object with a mouse that is used with the preferred hand. One advantage of this metaphor is that it combines tool selection and tool activation in a single operation.

A notable two-handed user interface is the T3, introduced by Kurtenbach *et al.* [1997]. They presented a GUI paradigm that is based on tablets, two hands, and transparency, hence the name T3. This paradigm was used in a sophisticated drawing application. The tools are controlled with two Wacom tablets that both have a puck that senses rotation in addition to their position.

Hinckley *et al.* [1997a; 1997b; 1998a] have developed a bimanual interface in which a doll is used to control neurosurgical visualization. They found that two hands provide more than just time savings over one-handed manipulation. Two hands together provide sufficient perceptual cues to form a frame of reference that is independent of visual feedback.

Multimodal user interfaces based on two-handed gestures and speech input were mentioned in the previous chapter. An early prototype was implemented by Bolt and Herranz [1992]. It also detected gaze direction using an eye tracker.

Commercial systems that use two-handed input have become available lately. SmartScene [MultiGen, 1999] uses two datagloves and is used for virtual prototyping. StudioPaint [Alias|Wavefront, 1999] is a digital sketching and painting system that is based on the T3 metaphor [Kurtenbach *et al.*, 1997]. It is probable that other systems will emerge when the benefits of multimodal user interfaces are realized and operating system support for handling parallel input devices becomes standard with the introduction of the universal serial bus [USB, 1999] that supports up to 127 concurrent devices.

3.4. Potential benefits of two-handed interfaces

Two-handed interfaces have potential benefits over one-handed interfaces. As a special case of multimodal user interfaces, many of the benefits of multimodality are also present in two-handed interfaces. The main difference to the general multimodal interfaces is that both devices are manual and thus do not provide redundant information that many multimodal systems use to disambiguate recognition errors. However, since the devices are manual, recognition errors are minimal in any case. Buxton and Myers [1986] found that two-handed interfaces are natural and efficient when they are used in tasks that are suitable for them. Zhai *et al.* [1997] found a clear advantage for two-handed input device setup compared to one-handed setup because the interface was less moded.

3.4.1. Two hands are natural

If we think of our everyday tasks, we use our both hands frequently to assist each other in performing several tasks, for example eating with a fork and a knife, tying one's shoelaces or driving a car. In contrast, we use only one mouse to control the computer.

Buxton and Myers [1986] experimented with using both hands to accomplish a combined resizing and repositioning task when manipulating objects in a drawing program. They trained the users so that they used only one pointing device at a time. However, 6 of 14 users used both hands in parallel from their first trial even though this was not suggested by the researchers. In total, two hands were used in parallel during 40,9% of total time in the trial. The amount of parallel use shows that using both hands in parallel is natural for the users. It was also efficient: the time spent in tasks correlated with the parallel use of both hands.

3.4.2. Two hands are efficient

Buxton and Myers [1986] also had a second task in which the goal was to find out whether using both hands would gain advantage in speed. The task was to select words in a word processor and scroll the page to find the next word. Normally both of these operations, navigation and selecting, are done with the same pointing device. In this experiment the preferred hand was assigned to selecting and the non-preferred to scrolling.

The results showed that with experienced users, a group that used both hands was 15% faster than a group that used only one hand. With beginners the

difference was even greater: the group that used both hands was 25% faster than the group that only used one hand. The difference between the novice and experienced users was smaller with the two-handed interface. With the one-handed interface, the experienced users were about 85% faster than the novices were, but with the two-handed interface the difference dropped to 32%. In addition, the novices that used the two-handed interface were only 12% slower than the experienced users that used the one-handed interface, and this difference was not statistically significant.

Dividing navigation and selecting to two hands made all the users perform better. However, in this task the better performance was not due to the parallel action of both hands, but a manual benefit of decreased movement time that was needed to switch between the operations in the one-handed interface.

3.4.3. Two hands are less moded

Zhai *et al.* [1997] repeated the second task that Buxton and Myers [1986] used. They applied the same navigate-and-select strategy to Web browsing. The task was to browse through the Web pages as fast as possible. The pages were designed so that selecting a link always required scrolling.

The goal of the experiment was to compare a standard mouse, a mouse with a wheel (Microsoft WheelMouse™), a mouse with an isometric joystick (IBM TrackPoint III™) and a two-handed mouse and isometric joystick setup. The standard mouse provided only one input, the others provided two input feeds, but only the two-handed mouse and joystick setup used both hands and resembled the input setup in the earlier work [Buxton and Myers, 1986].

The results showed that the two-handed technique was the fastest, but the mouse with an isometric joystick was almost as fast. However, the mouse with a wheel was even slower than the standard mouse. It seems that even if the use of the wheel is intuitive, cognitive demand increased so much that speed slowed down.

Subjective ratings supported the ordering of devices: the mouse with an isometric joystick and two-handed interface were clearly the most preferred, the standard mouse got a slightly positive rating, but the mouse with a wheel got a clearly negative rating. Zhai *et al.* state that the devices in two hands do not need to be same kind of devices in order to perform better and have better subjective preference.

3.5. Guidelines for two-handed interaction

A badly designed two-handed interface can be even worse than a normal one-handed interface. More degrees of freedom require more responsibility from the system designer. Guidelines for two-handed interaction have thus been proposed to avoid common mistakes.

Chatty [1994] suggests that there should not be a task in the interface that demands using both hands. This requirement is based on the fact that some people may not be able to use their second hand due to disability or another task, as holding the phone or a cup of coffee. However, if there are methods that use one or two hands for the same task, these methods should behave in a similar manner so that the user will not be confused. Chatty also proposes that natural two-handed interaction is accomplished by applying real-world metaphors and extending them to work bimanually.

Chatty [1994] presents a comprehensive set of guidelines for two-handed input. The classification of multimodal user interfaces [Nigay and Coutaz, 1993] provides a good framework for presenting these guidelines, in which the more complex interaction technique builds on the simpler techniques. The following discussion is divided in three classes of two-handed interfaces: independent interaction, parallel interaction, and combined interaction.

3.5.1. Independent interaction

A simple way to extend one-handed interfaces is to add a second pointing device that is used in the same way as the first. This technique is only useful when both devices control their own cursors. The addition of the second pointing device can speed up interaction in a way similar to the selecting and scrolling task by Buxton and Myers [1986]. These systems may have a manual benefit because they require less cursor movement when two actions take place sequentially. This kind of two-handed interface is also easy to implement, provided that the system supports two pointing devices.

The second pointing device could also be used for a more complex task, such as drawing pictures or writing text. However, the precision of the non-preferred hand is worse and the error-proneness greater than those of the preferred hand. These disadvantages of the non-preferred hand can partly be compensated by making the objects large or making the cursor large. The non-preferred hand may still be slow, but tasks that require less granularity are now possible for it. However, the real benefits of two-handed interaction are realized when the two tasks can be performed in parallel.

3.5.2. Parallel interaction

It is natural for humans to use both hands in parallel. Even though we have not been taught to do so, we often use our non-preferred hand in secondary tasks like reaching for a tool that is used with the preferred hand. If two-handed user interfaces forced a sequential control, their use would be very constrained: it is not natural for our hands to wait for one another doing nothing when we perform common tasks in the real world. Thus parallel interaction should be a required property of two-handed interfaces.

Both hands could also be used in independent but parallel tasks of the same importance. This kind of interaction is related to specific applications in which the user is motivated to learn this somewhat more difficult interface. Obvious examples are flight simulation and computer games. There may also be cases in which expert users of computing systems would rather learn a little more difficult way of control if that would result in better efficiency. Normal computer users will experience two-handed interfaces as natural and efficient when the actions of both hands are related and combined to achieve a common goal.

3.5.3. Combined interaction

In the real world we often use both hands to achieve a common goal: for example, the non-preferred hand holds a nail when the preferred hand uses a hammer. Another natural use for the non-preferred hand is to support the preferred hand and provide additional strength in the movement. One example of this is playing golf.

In traditional user interfaces the second hand is often replaced with a kind of magic. For example, when we move one end of a line in a drawing editor, the other is held with an invisible, magic hand. Chatty proposes that this magic should be disabled when two hands are at work. The non-preferred hand could hold the other end of the line if it should stay where it is. If the non-preferred hand were not used, the preferred hand would drag the whole line. He names this kind of interaction style as “hold-and-pull”. A parallel, but not combined interaction technique would be to use the non-preferred hand to move one end of the line at the same time as the preferred hand moves the other.

Another kind of combined two-handed technique can be used in critical operations. For example, if shutting down the system required selecting a button with both pointing devices, it is not probable that this would happen by accident. To be accepted, this kind of dual clicking would require the other click

to occur within a certain time, as is the case in double clicking. However, this time interval should probably be longer than with a double-click. This kind of confirmation technique has actually been used for years in common cassette and video recorders that often require the play button to be pressed at the same time as the record button.

Chatty's guidelines are very general and do not give specific recommendations on two-handed interaction techniques for a given task. Two-handed interaction is a much smaller domain than more general multimodal interaction. However, the questions presented in the previous chapter are valid for two-handed interfaces as well: we do not have a thorough understanding of when and where to use two-handed interfaces, which input devices to use in a given system, and which interaction techniques to apply. Many research papers that were presented in this chapter give suggestions of tasks that benefit from two-handed interaction. The publications that belong to this dissertation aim partly at advancing this knowledge.

3.6. Summary

This chapter summarized earlier research in two-handed interaction, and presented psychological theories and empirical results that help in designing two-handed user interfaces. There are many tasks in which two-handed interfaces have been found beneficial, such as resizing an object while it is moved [Buxton and Myers, 1986], understanding three-dimensional manipulation [Krueger, 1983; Hinckley *et al.*, 1997ab, 1998a] and using the second hand for a less detailed task [Kabbash *et al.*, 1994]. Chatty [1994] provides guidelines for two-handed interaction, but there is not a general understanding of the tasks that would benefit from two-handed input. Empirical results and psychological theories help in deciding when to consider these interfaces. Because we do not have a thorough understanding of two-handed interaction, it is essential that the systems are evaluated empirically when these interfaces are designed and implemented.

Two-handed interfaces have been found to be faster and easier to use than conventional interfaces that are based on one mouse and a keyboard. However, these benefits are only realized if two-handed interfaces are designed well. A generic result is that the actions of both hands need to be related to the same task. If they are not, manual control requires much more cognitive resources, and will be more difficult and most likely slower than in normal one-handed interfaces.

4. Multimodal interaction techniques

This chapter summarizes the research papers on multimodal interaction techniques that belong to this dissertation. Often the concept of an interaction technique is assumed to be self-evident. Most researchers who discuss these techniques do not define the concept at all. A definition by Robert Jacob [1993] is by no means exhaustive, but in my opinion it captures the essence of the concept of an interaction technique well:

“An *interaction technique* represents an abstraction of some common class of interactive task, for example, choosing one of several objects shown on a display screen. Research in this area studies the primitive elements of human-computer dialogues, which apply across a wide variety of individual applications.”

Foley *et al.* [1990] have described an interaction technique as a way of using a physical input device to perform a generic task in a human-computer dialogue. Both of these definitions state that an interaction technique is a way to carry out an interactive task. The interaction techniques can be used as building blocks for new kinds of user interfaces.

The papers present three different cases of multimodal interaction techniques:

1. *The alignment stick*: an interaction technique for object alignment in drawing programs and diagram editors. This tool is based on direct manipulation and two-handed input. The goal was to build a tool that would be both intuitive and efficient to use.
2. *An alternative way of drawing*: a group of tools that offer an alternative for manipulating splines or other mathematical structures in drawing programs. This new way of drawing offers a different point of view in creating object-oriented drawings.
3. *Touchscreen interaction techniques combined with speech recognition*: several selection techniques were implemented in a multimodal kiosk prototype that is used with touchscreen and speech recognition. Three techniques used pressure detection, one was time-based, and one was based on direct manipulation.

Two research prototypes were implemented to carry out research in these subjects. *R2-Draw* is an object-oriented drawing program that implements the alignment stick and our alternative way of drawing. It is controlled with two-handed input. *Touch'n'Speak* is a multimodal kiosk prototype that uses many modalities for both input and output. Thus, a large part of this dissertation was constructive research.

Several empirical studies were performed to evaluate the usefulness of these new interaction techniques. In total, almost 100 users performed typically a one-hour session in evaluating one of the interaction techniques. The details of each study are explained in the following papers.

4.1. A new direct manipulation technique for aligning objects in drawing programs

The first paper [Raisamo and R  ih  , 1996] describes the iterative design of a new direct manipulation tool for alignment tasks. This tool, *the alignment stick*, replaces the alignment commands that the user usually selects from pull-down menus or tool palettes. The design process was constructive and involved informal and empirical testing in different stages of the tool development.

The design of the tool started with a careful consideration of the metaphor for the tool. We selected the ruler as the metaphor for the alignment tool. If the user pushes objects with this tool, the objects are aligned with a single gesture, and the alignment is animated in real time when the tool is used. Figure 12 shows how three ovals are aligned by their bottom sides when the alignment stick is moved upwards.

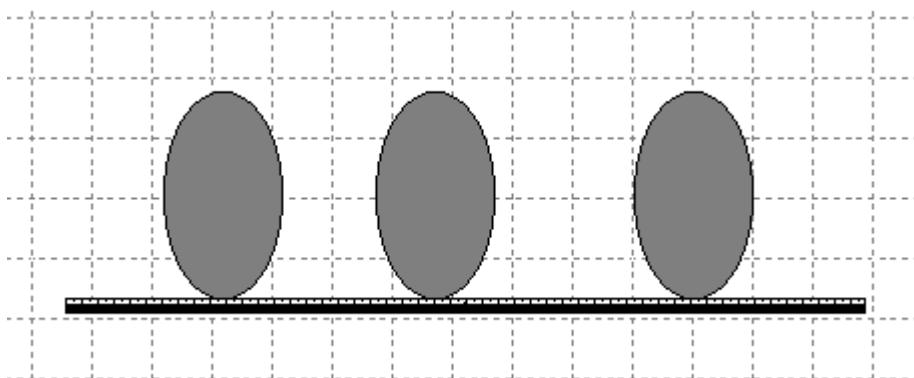


Figure 12. Aligning three ovals by their bottom sides with the alignment stick.

The user can select any combination of *alignment points* that are the active points that the stick grabs during the alignment. This way the stick can also align the objects by their center points, or can even do more complex tasks as aligning the top of one object to the center of the others (Figure 13).

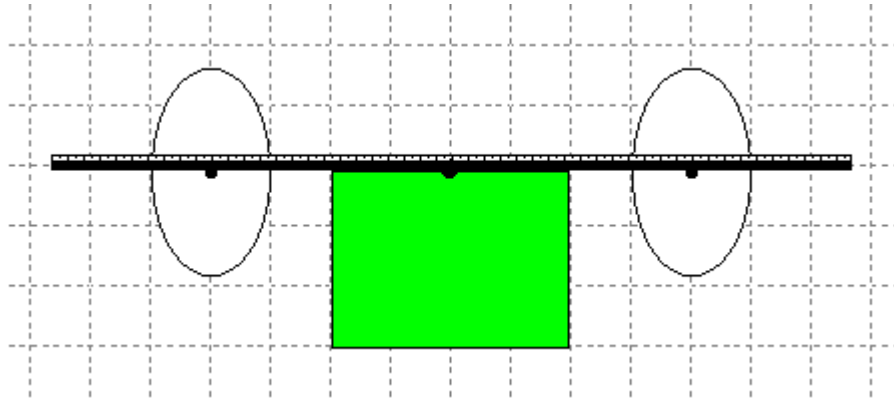


Figure 13. Using the alignment stick to align the top of the rectangle to the center of the ovals.

The alignment stick is controlled with a two-handed user interface, in which a mouse is used to control the position of the tool with the preferred hand and a large trackball is used to control the length of the tool or rotation angle of the tool with the non-preferred hand. Because these properties are related to the same tool, and can be changed in parallel, they are a case of combined interaction [Chatty, 1994] which is the class of tasks that is the most suitable for two-handed interfaces.

The two-handed control of our interaction techniques [Raisamo and R  ih  , 1996; Raisamo, 1999a; 1999b] has been designed keeping in mind the psychological studies and empirical results presented above. They are another example of using two hands in an integrated manner to control interactive tools. Both hands provide continuous input feeds that are combined in a single action.

4.2. Design and evaluation of the alignment stick

The main topic of the second paper [Raisamo and R  ih  , 1999] is an empirical comparison between command-based alignment tools and the alignment stick. We compared alignment menus, tool palettes and the alignment stick in different alignment tasks. Twenty-two users took part in the main experiment and performed several tasks with all three alignment methods. This study showed that we were able to create a tool that is intuitive for the users.

However, the design was not without problems. When the alignment points were needed with overlapping objects, the tool became hard to use. This was because the user needed to switch between different tools and arrange the objects in depth order to be able to select alignment points of all needed objects.

We noticed that the alignment stick was especially useful with non-overlapping objects and in tasks that are needed in all diagram editors. The most complex task that involved overlapping objects did not suit well for the evaluated alignment stick. This kind of task would not be common in diagram editors, but is a basic task in object-oriented drawing programs.

In general, the tool palette performed best and was preferred by the users. The menu was the least preferred alignment method and produced most orientation errors when the user selected an alignment command. An interesting result was that there were no statistically significant differences in execution time between the palette and the menu. In addition, there were no statistically significant differences in the accuracy of the three alignment methods.

Robbins *et al.* [1999] also compared a tool similar to our alignment stick to standard alignment commands. They mentioned that the majority of the subjects in a small study found the tool more "natural". They also found that the tool required less mouse movement and dragging than the standard tools.

4.3. An alternative way of drawing

An alternative way of drawing was developed by extending the stick metaphor that was used in the alignment stick. The third paper [Raisamo, 1999a] explains a set of direct manipulation tools that can be used in creative drawing. The goal was to offer an alternative to manipulating splines and other mathematical structures that are not at all clear for an average user as noted in previous work [Baudel, 1994; Fowler, 1992; Gross and Do, 1996; Qin and Terzopoulos, 1995; Terzopoulos *et al.*, 1987].

The new tools were the carving stick, the shrinking stick, the cutting stick and the rotating stick. The tools can be used to sculpt objects in two dimensions. The major difference between this alternative way of drawing and common drawing programs is the interaction style. With current object-oriented programs, the user usually builds drawings from many separate pieces. With this alternative way of drawing, one or more large blocks are first drawn on screen, after which the user starts sculpting them. Of course, the user can make as many objects as he or she wants at any time, but still this paradigm shifts

focus from composing a drawing from small pieces to manually sculpting different forms.

In a proof-of-concept evaluation, six users tested the new tools and produced some drawings. The users succeeded in drawing a variety of different forms within a very short time (15 - 30 minutes). All the tools were controlled in the same way as the alignment stick, and were thus based on two-handed interaction. The controls had been fine-tuned after our initial paper [Raisamo and R  ih  , 1996].

The controls are presented in Figure 14 using the Petri net model by Hinckley *et al.* [1998b], see section 3.2. As can be seen, when the left mouse button is pressed the tools are always in one of the combined input modes, $S2_{np}$ or $R2_{np}$, and this mode is selected with the middle mouse button. S denotes changing the size of the tool and R denotes rotating the tool. The right mouse button that allows emulating the trackball with the mouse has been left out of this model, since it is not used in normal conditions.

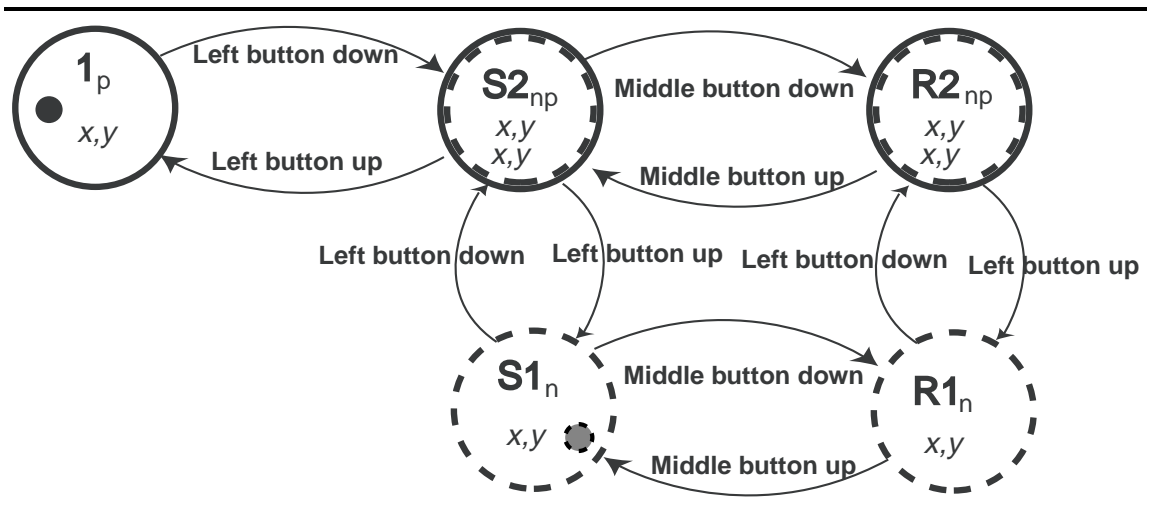


Figure 14. A Petri net model of the revised controls of all stick tools.

4.4. An empirical study of how to use input devices to control tools in drawing programs

The fourth paper [Raisamo, 1999b] presents an empirical evaluation on using five different controlling methods for stick-shaped direct manipulation tools. The different methods were (1) a mouse, (2) a mouse and a keyboard, (3) a mouse and a trackball, (4) a touchpad and a keyboard, and (5) a touchscreen and a keyboard. In total, twenty-nine users tried all five different input methods and compared them by filling in an evaluation form. Quantitative information

was collected automatically in the research prototype. We recorded the number of alignment operations and time for each operation. We also recorded the exact positions of three sample rectangles that the users were asked to align to a specific position at the end of each evaluation session.

The subjective ratings showed a clear preference for the two-handed mouse and trackball method. The other mouse-controlled methods were rated the next best. The touchscreen and touchpad did not receive good ratings because they were not as easy to control as the mouse-based methods. The users did notice that the touchscreen was the most direct input method, but would also have expected dragging to be as easy and accurate as it is with the mouse.

There were no statistically significant differences in the accuracy of input methods in the measured positioning task when we compared the final positions. The amount of operations was about the same in all three mouse-based methods, over twice as great with the touchpad method, and about four times as great with the touchscreen method. However, the average length of a single operation was clearly the shortest with the touchscreen setup. The users reported accidental loosening of touch when performing touch operations, which caused the short average length and the large number of operations. The evaluation gave support for two-handed interaction, but did not support using a touchscreen to control direct manipulation tools.

4.5. A multimodal user interface for public information kiosks

The fifth paper [Raisamo, 1998] presents *Touch'n'Speak*, a *multimodal information kiosk framework* that responds to both touch and speech input. Touch input is used in several ways: the user can touch large on-screen buttons, touchable lists, or make a selection in a picture by using different touch-based area selection gestures. The user has always an option to speak the same alternatives. In addition, speech is used to support touch in area selection techniques.

Figure 15 presents a high-level model of the Touch'n'Speak system. This model is a simplified version of the general model in Figure 4 (section 2.2). The input modalities have been divided in three categories: touch position, touch pressure, and speech input. The output media are speech, graphics, text and sound, or auditive and visual modalities. I have left “plan recognition and generation” out of the interaction management section, since Touch'n'Speak is quite simple a system and does not apply any artificial intelligence algorithms.

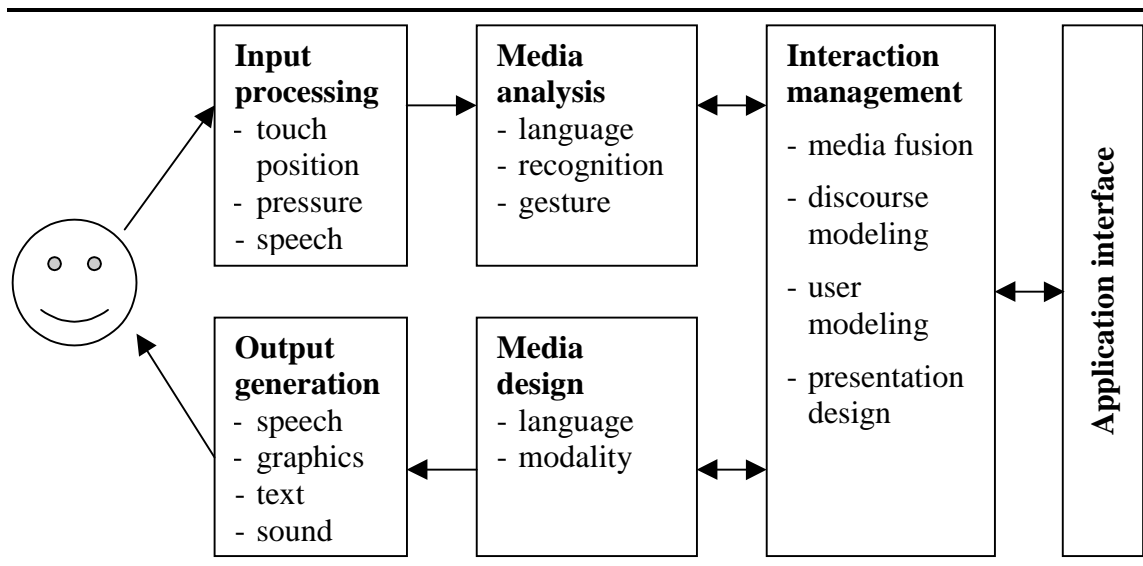


Figure 15. A high-level model of the Touch'n'Speak system.

The application interface is in practice the specialization interface of this object-oriented white-box framework. This framework was used to implement a multimodal restaurant information system that has information on restaurants that are situated in Cambridge, Massachusetts. The structure of the framework is explained in the paper using this example application.

4.6. Evaluating touch-based selection techniques in a public information kiosk

In the last paper [Raisamo, 1999c] I report on an empirical study in which 23 users tested our multimodal kiosk prototype [Raisamo, 1998]. I found that pressure-sensitive techniques offer a possible alternative to the other techniques that were based on time and direct manipulation. Still, pressure-sensitive techniques are similar to two-handed techniques in that they require very careful designing. It may not be obvious for the user that the screen detects pressure in addition to basic touch if that is not visible. People may also have difficulties in controlling the touch pressure accurately.

The time-based selection technique was the most intuitive for the users and received the best subjective ratings. The direct manipulation technique was rated the second, but it required the longest time to understand how it works. One pressure-based technique was almost as good as the direct manipulation technique, but two others were rated clearly the worst.

The users strongly preferred multimodal interaction to interacting with only one modality. Touch was the more preferred modality if used as the only input

modality. This was partly because speech recognition did not work well with all the users that were non-native English speakers. Still, the users had a very positive attitude towards speech. Without exception they commented in a related interview that they would like to use speech and that it would be useful if it worked well.

There are not many published experiments on the use of public information kiosks. In an earlier study on information kiosks Salomon [1990] analyzed CHI'89 InfoBooth kiosks that were built using an Apple Macintosh computer, Hypercard software and a mouse as the only input device. She describes an incremental design process that was used to build the system. Her results can be applied to touch-based interaction since the interaction style was similar to ours: a mouse was used as the only input device and it was used to select items on screen.

4.7. Summary

The two research prototypes that were constructed represent the two different categories of multimodal systems that were presented in section 2.1. *R2-Draw* clearly follows the computer as a tool paradigm and uses two-handed interaction to enhance direct manipulation with novel drawing tools. In contrast, *Touch'n'Speak* is a conversational system that is based on computer as a dialogue partner paradigm. While *R2-Draw* is a pure example of its paradigm requiring the user to initiate all operations, *Touch'n'Speak* also provides the user with direct manipulation techniques in addition to the conversational interface. This makes it a mixed-initiative system that has lately been a common research approach in many speech recognition systems.

The research papers presented in this chapter suggest new interaction techniques for two-handed input in desktop systems and multimodal input in public information kiosks. The alignment stick has been found to perform well in its basic use. The alternative way of drawing frees the user from manipulating control points in object-oriented drawing programs. While the main goals of these interaction techniques were accomplished, there is still room for future improvements.

Our information kiosk framework allows speech and touch input. The new pressure-sensitive selection techniques did not yet perform as well as the time-based and direct manipulation selection technique, but pressure-sensing is still an interesting input channel and may result in efficient interaction techniques in later studies. Now that we have determined the intuitiveness of these selection

techniques, the proper help information can be added and changes in their speed and other parameters can be done.

Different kinds of contributions are present in the discussed work. First, new interaction techniques were presented for certain tasks. Second, two-handed and multimodal interaction were successfully used to control these interaction techniques. Third, we found out how the selected input devices performed with these interaction techniques. Fourth, we got other empirical results on the use of the two constructed prototypes. In respect to my urging in section 2.6 that we need to understand multimodal interaction better, we can see that my results only partly answer the last three questions that I presented. Even though I have taken earlier empirical results and psychological theories into account, it is not possible for me to claim that the best modalities, input devices, and interaction techniques were found for the given systems and tasks.

In the next chapter I present ways to further develop the interaction techniques that were presented here, and other development that is possible when continuing this line of research. Some of that work is already under development.

5. Further research

In this chapter I discuss the consequences and new developments of the work that was presented in the research papers. I propose possible applications for the presented interaction techniques and suggest future improvements that could be made in each group of techniques.

5.1. Alignment tools

Chatty [1994] suggested that there might be cases in which the expert users of computing systems would rather learn a little more difficult way of control if that would result in better efficiency. Our results in the second paper [Raisamo and R  ih  , 1999] seem to support this since the experts appreciated most the advanced functionality of the alignment stick that required controlling the alignment points.

The trackball was found to perform well when used with the non-preferred hand. In the second paper [Raisamo and R  ih  , 1999] the users gave high subjective ratings for using the trackball, and in the fourth paper [Raisamo, 1999b] it was preferred to other input device configurations. The two-handed mouse and trackball setup was also found to perform among the best. Bier *et al.* [1993] have also used the trackball to control toolglasses and magic lenses with the non-dominant hand.

Still, there is need for a further study in which the users could use the alignment tools to align parts of a large diagram in a diagram editor. This kind of interaction incorporates zooming and panning as two of the main tasks since a large diagram can not be presented clearly in a small screen. Zooming and panning are especially suitable for the non-preferred hand, and could be implemented by using the trackball buttons that were not used. However, it is possible that the increased complexity requires a separate zooming and panning tool.

The alignment stick was found to be intuitive in its basic functionality in [Raisamo and R  ih  , 1999]. There are still some issues in its design that can be further developed and other similar tools can be constructed. They are discussed in this section.

5.1.1. Alternative shapes for alignment tools

The alignment stick is an example of a group of possible alignment tools. Different forms of tools can be envisioned for different tasks. Figure 16 shows a curve-shaped alignment tool that is present in our new Java-based implementation that is currently under development. One example of using the curve-shaped alignment tool in an ER diagram editor is to align the properties (ovals) around an entity (rectangle) in a nice round shape.

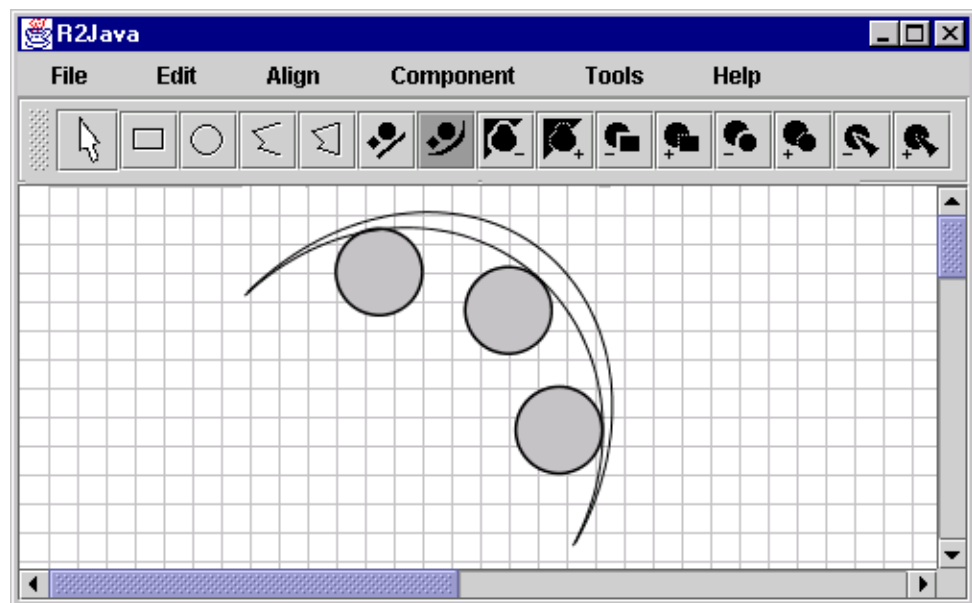


Figure 16. A curve-shaped alignment tool in a new Java implementation.

5.1.2. Making alignment points and grouping more intuitive

The largest problem with the alignment stick was found to be the need to select alignment points and group the objects in more complex alignment tasks [Raisamo and R ih a, 1999]. Even though these features allow added functionality to the present drawing systems, they also made the tool less intuitive for the users.

Alignment points need to be selected in some way to allow other alignments than aligning by the border. Grouping is needed when some part of the drawing already is in the intended alignment. It is also needed when a group of objects is aligned to other objects since a group has its own alignment points. However, it should be noted that grouping is also needed with the conventional alignment commands when they are used in similar alignment tasks.

One way to make grouping easier would be to add intelligence in the interface so that grouping could be done automatically in some situations. Then the user would not need to switch to the pointer tool and group the objects, but the objects would be instantly grouped. A promising grouping technique can be to create automatic grouping constraints when the tool is used to align objects. For example, if some objects are aligned by their center points, it is probable that the users want to keep them that way. However, this automatic grouping should have a different status than the explicit grouping command. Automatic groups should be presented with a different screen presentation, and it should be possible to override them easily if the user wants to align this inductive group in a new way. I got the idea for this kind of grouping from a paper by Olsen and Deng [1996] in which they applied inductive groups in a drawing package which was used to draw diagrams.

The handling of overlapping objects has been made easier in the new Java prototype. All opaque objects are made transparent during the alignment operations. Therefore, if a smaller object is accidentally hidden behind a larger one, it can still be seen and manipulated. We have also made it easier to change the depth order of overlapping objects. If the pointer tool or the alignment point tool is selected, the middle mouse button cycles through the overlapping objects. This feature should make setting alignment points easier since there is no need to switch to the pointer tool to get a hidden object on top of another.

5.1.3. Bridging the virtual and physical worlds

A growing trend in human-computer interaction is bridging the virtual and physical worlds. There are promising research opportunities for using more physical tools to control the virtual tools that are drawn on screen.

We could build a more concrete user interface in which the tools would really be physical tools and not indirectly controlled virtual representations of physical tools. Fitzmaurice and Buxton [1997] have used touch tablets to implement such tools. Until now our two-handed tools have been controlled indirectly with two pointing devices – still based on direct manipulation. Even though the tools follow the principles laid by Shneiderman [1982; 1983] we can envision more direct ways to control these tools. We could use a real ruler that the system would detect with some sensor or machine vision technology. Electrical field sensing [Zimmerman *et al.*, 1995] is another possible option, but I would prefer other solutions, because not all people are likely to accept electric current – even a very weak one – to be conducted through them. This approach

follows the principles that Fitzmaurice *et al.* [1995] suggested for graspable user interfaces.

Using physical tools makes the interface more direct, but this approach has its own problems. One design decision is the arrangement of the display: should the screen space and input space be the same or should they be separated. My results [Raisamo, 1999b] suggest that the users would have liked the directness of the touchscreen had it been less error-prone. Better ergonomics and less hand fatigue may be gained by using touchscreens that are mounted in horizontal or tilted LCD displays. With all techniques that have the screen space and input space merged there are problems when the tool or a hand occludes a part of the screen. However, if the input device and screen are separated, some amount of directness is lost even if the input device, for example a graphics tablet, is used in an absolute mode, which means mapping the tablet surface directly to the screen.

5.2. Two-dimensional sculpting metaphor

Two-dimensional sculpting metaphor was found to be plausible in a user evaluation [Raisamo, 1999a]. The new direct manipulation tools presented in this dissertation were all straight sticks. We have envisioned other forms of tools that would be useful: curved tools, sharp-ended or round-ended tools and free-form objects as sculpting tools.

Figure 17 shows another screenshot of the new Java implementation of R2-Draw. There a cheese-shaped object has been selected as a sculpting tool that is rotated and is going to be used to cut the mushroom. There are many different sculpting tools in the tool palette of which the free-form sculpting tool is currently selected. With the free-form sculpting tool, any object on screen can be selected as the sculpting tool that can be resized and rotated. The tools with a minus sign cut the objects like the earlier carving stick [Raisamo, 1999a]. The tools with a plus sign add to the objects as if the sculptor added clay in the sculpting process. Different forms of tools have been provided to speed up using the most common tool shapes.

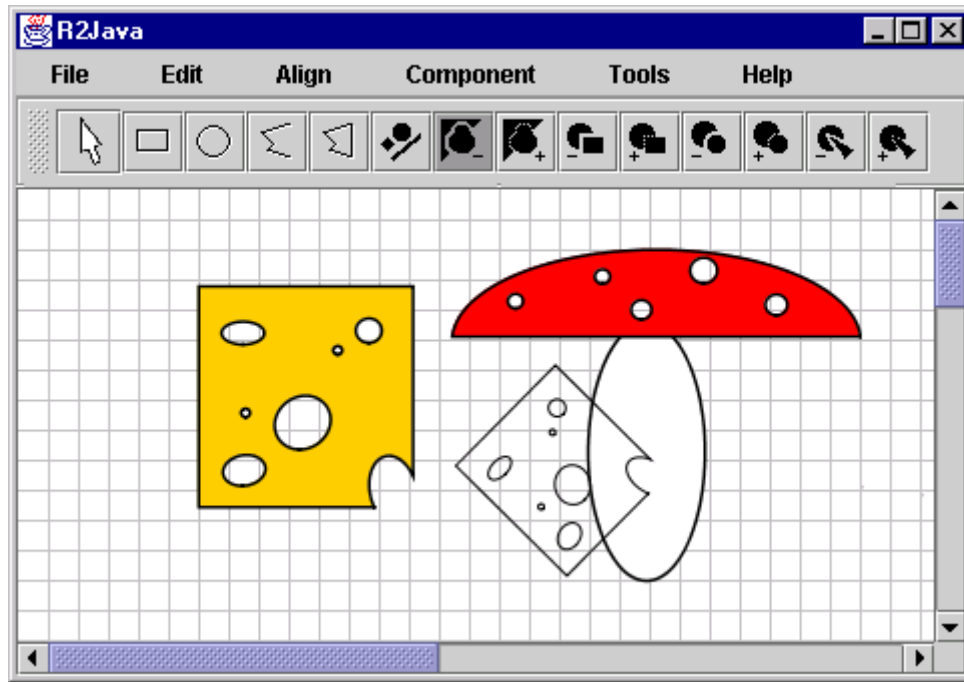


Figure 17. The new Java implementation allows arbitrary carving tools.

5.3. Touchscreen selection techniques

The different touchscreen selection techniques that were presented in [Raisamo, 1999c] can be applied in public kiosks. In addition, area selections are often needed in conventional desktop environments, and these techniques can also be used in them if the computers are equipped with touchscreens. However, the fatigue that a vertical touchscreen causes in extensive use is a problem unless the screen is mounted in a tilted orientation.

Touch-based techniques can also be applied to other touch-based input devices, like touchpads and touch tablets. These devices do not have display capabilities, but the selection techniques do not necessarily require that. Many touchpads and touch tablets recognize touch pressure with some accuracy. There are many potential applications in using touch pressure to something other than just defining the width of the pen in drawing programs, as in Wacom's PenTools plug-in drawing tools [Wacom, 1999].

Several suggestions for selection techniques were discovered during the experiment that was described in [Raisamo, 1999c]. A two-handed selection technique was mentioned in the paper. We would have implemented it before the experiment if the hardware had allowed for it. The other three may be taken into account in the future development of our system. They are multi-fingered

selection techniques, tapping selection techniques, and encircling techniques that are described next.

Closely resembling two-handed selection techniques, multi-fingered selection techniques also require that the screen detects many simultaneous touch points. Some of our test users proposed a selection technique in which each finger could point at a different area that needs to be selected at the same time as the others. The selection area could not be very large since its size is limited by the size of the user's hand. This technique sounds interesting and can be tested when touchscreens and their drivers will detect many touch points.

As many as nine users tried repetitive poking of the map. This hindered their understanding on the tested selection techniques since all of them required continuous touching. The main argument for using tapping was that it's the way the other parts of the screen, like buttons and list box items, are selected.

One simple idea that has been suggested is encircling an area by drawing its boundaries with a finger. This should be intuitive, but it requires longitudinal and continuous pressing on screen, and is not convenient for many users as we found in another user evaluation [Raisamo, 1999b]. The screen that we used requires so much pressure that it is not very nice to press the finger on screen and move it for a longer time.

Lastly, different areas may need to be selected at the same time. Now that we have tested each selection technique in making single selections, a mechanism needs to be added to combine many selections. This mechanism can be quite trivial such as selecting several objects with the pointer tool when the shift button is pressed. Speech commands could be used in this task, but they should not be the only option, since sometimes speech recognition does not work reliably.

5.4. Enhanced multimodality

As noted in Chapter 2, there is an abundance of possibilities to build multimodal interfaces. The two-handed drawing program prototype and the multimodal kiosk prototype make use of some of them. Both systems can be augmented with new interaction modalities that may make them perform better and may allow the system to figure out when the user needs guidance with some operations. Here I discuss some modalities that could be added in these systems.

5.4.1. Feet in desktop interfaces

The two-handed interfaces can be enhanced in many ways. One way to increase the bandwidth between the user and the computer is to make use of feet. Clearly, the feet are not as accurate and sophisticated as our hands, but a secondary function may well be assigned to a foot. They can act in a way as our third and fourth hand which are even less accurate than the non-preferred hand and twice as slow as the hands [Hoffmann, 1991]. Most people can drive a car and use both hands and feet when driving. This is possible because the feet have very simple and constrained tasks. Something like pressing a gas pedal can well be integrated into multimodal user interfaces.

Feet have already been used in specialized user interfaces like piloting a military aircraft. Taylor [1989] discusses practical issues that need to be taken into account when integrating voice, visual and manual transactions. Many highly interactive games make an efficient use of feet when they simulate car racing or piloting a military aircraft. These games usually try to mimic their real counterparts as closely as possible, which is a strong reason for utilizing feet in their controls. These simulators are, however, closer to virtual reality systems than multimodal user interfaces.

5.4.2. Machine vision techniques

Information kiosks may make use of many different input modalities. The fifth paper [Raisamo, 1998] presented a touch and speech based kiosk prototype. Additional input modalities could be integrated in the system, such as machine vision and eye tracking. However, they share some of the same problems as speech in kiosk systems. The users can not be effectively restrained to stay in a certain place and to look at the screen all the time. Moreover, as reported in [Christian and Avery, 1998], the problems with crowds and erratic lighting conditions that were encountered in deploying the Smart Kiosk prototype in SIGGRAPH '97 exhibition suggest that machine vision may not work well in many practical situations.

It is interesting that the processing power of many current workstations is enough for using some basic machine vision techniques. Recently, many inexpensive and still powerful video cameras have appeared in the market to support teleconferencing. With appropriate software these cameras can be applied to sense the user in desktop environments and provide complementary input for other input devices. The real advantage of using machine vision is that it is totally seamless: it does not require anything from the user. The system can

monitor the user and interpret input that can disambiguate other inputs or give hints of the user's mental state. For example, if a user is frustrated, we should probably offer some kind of help in using the system. Machine vision can also be used to detect whether the user is present, and to adapt the interface accordingly.

5.5. Summary

We have just begun exploring the great possibilities that multimodal user interfaces have to offer to the current way of computing. This chapter introduced enhancements to the published interaction techniques along with ways to use other modalities to provide additional input in these interfaces. New tools and tool shapes have been described for other tasks that are similar to tasks in the earlier work [Raisamo and R  ih  , 1996; Raisamo, 1999a]. The Java-based prototype drawing program will later be used in a longitudinal study of drawing with the proposed alternative way of drawing.

The techniques are successful only if the users accept them and make use of them. After receiving the encouraging subjective ratings of the new drawing tools and making the minor improvements in the new Java prototype we have released it on the Web [Raisamo *et al.*, 1999]. This gives the users an opportunity to try it out and give their comments. This large-scale evaluation will presumably take plenty of time, but should produce long-term results on the use and acceptance of the new techniques. It should be possible to collect some log data and subjective evaluations of the tools when voluntary Internet users try them out.

6. Conclusions

This work presented a survey of literature on multimodal interaction and two-handed interaction. Multimodal user interfaces can be seen as the next user interface paradigm that current powerful computers and better understanding of human-computer interaction have enabled. Two-handed interfaces as a special case of multimodal interfaces allow the user more degrees of freedom when controlling the computer. However, it is very important that different modalities and both hands are used in a way that is natural for the users. Each modality has its own strengths and weaknesses. We currently know quite little of which modalities to combine, when to build multimodal user interfaces and how to use the selected modalities to design an interaction technique for a given task. There is plenty of inter-disciplinary research that needs to be done by specialists of human-computer interaction and psychology to resolve these questions.

The two main contributions of this work are the description of constructive research that was carried out to design and implement multimodal interaction techniques, and the evaluation of the new interaction techniques. The presented techniques were implemented in two research prototypes, an object-oriented drawing program R2-Draw and a public information kiosk framework Touch'n'Speak. Several multimodal interaction techniques were presented for aligning objects, sculpting drawings and making touch-based selections with spoken commands in public information kiosks. The new two-handed direct manipulation tools aimed at making certain operations in drawing programs more intuitive and yet efficient for the users. The touch-based and pressure-based area selection techniques were presented as a part of the multimodal information kiosk framework.

Four empirical studies of the new interaction techniques were presented. In one of the studies [Raisamo and R ih , 1999] we compared the alignment stick [Raisamo and R ih , 1996] to the conventional palette and menu based alignment commands. In the next study we evaluated the alternative way of drawing that is used to sculpt drawings in two dimensional drawing programs [Raisamo, 1999a]. Five different ways to control these tools were compared in another study [Raisamo, 1999b]. It provided evidence for our design decision to choose a two-handed interface that is controlled with a mouse and a large

trackball. The last evaluation [Raisamo, 1999c] compared five different ways to make area selections on touchscreens using time, pressure and direct manipulation. The evaluation was done using the multimodal public information kiosk prototype [Raisamo, 1998].

In Chapter 5 I discussed the practical consequences of the empirical studies and proposed improvements that can be made in the techniques that were presented. Some of these improvements were already under development and described in that chapter. Going beyond the limits of current user interfaces suggests many new research opportunities. Real-world tools could be used to bridge the physical and virtual worlds and make the graphical operations even more direct and tangible than they are with the best direct manipulation interfaces. Many recognition technologies, like speech recognition, machine vision, and eye tracking may be used to enhance public kiosks and desktop user interfaces.

Multimodal interaction is becoming more common as our understanding of its design grows better and new specialized interaction devices become inexpensive enough to allow mass-market production of multimodal user interfaces. Two-handed interfaces using two pointing devices are possible to implement with all current personal computers if the support for handling parallel input is implemented in software. With the recent introduction of Universal Serial Bus accessories and parallel input device architectures that are actively developed to enable highly interactive arcade games, the prerequisites for multimodal user interfaces are soon available in all modern personal computers. Then the software companies should become interested in leveraging the possibilities that multimodal user interfaces can offer for human-computer interaction. At that stage we should be able to provide guidelines for designing and implementing multimodal user interfaces. Otherwise, this promising opportunity of human-computer interaction may not be realized outside research laboratories.

References

- [Alias | Wavefront, 1999] Alias | Wavefront Inc., StudioPaint product home page. http://www.aw.sgi.com/design/products/consumer_products/studiopaint/ (checked 15 Nov 99)
- [Allport, 1980] D. A. Allport, Attention. In: *New Directions in Cognitive Psychology*, G. L. Glaxton (Ed.), Routledge and Kegan Paul, London, 1980.
- [AMODEUS, 1995] P. J. Barnard, N. O. Bernsen, J. Coutaz, J. Darzentas, G. Faconti, N. H. Hammond, M. D. Harrison, A. H. Jørgensen, J. Löwgren, A. Maclean, J. May, and R. M. Young, *The AMODEUS Project, Esprit Basic Research Action 7040 Final Report*, July 1995.
<http://www.mrc-cbu.cam.ac.uk/amodeus/d.html> (checked 15 Nov 99)
- [Annett *et al.*, 1979] J. Annett, M. Annett, P. T. W. Hudson and A. Turner, The control of movement in the preferred and non-preferred hands. *Quarterly Journal of Experimental Psychology*, 31, 641-652.
- [Baber, 1997] Christopher Baber, *Beyond the Desktop: designing and using interaction devices*. Computers and People Series, Academic Press, 1997.
- [Baddeley, 1986] A. D. Baddeley, *Working memory*. Oxford University Press, 1986.
- [Baecker and Buxton, 1987] Ronald M. Baecker and William A. S. Buxton, Case study D: the Star, the Lisa, and the Macintosh. In: *Readings in Human-Computer Interaction, A Multidisciplinary Approach*, R. M. Baecker and W. A. S. Buxton (Eds.), Morgan Kaufmann, 1987, 649-652.
- [Balakrishnan and Kurtenbach, 1999] Ravin Balakrishnan and Gordon Kurtenbach, Exploring bimanual camera control and object manipulation in 3D graphics interfaces. *Human Factors in Computing Systems, CHI '99 Conference Proceedings*, 1999, 56-63.
- [Baudel, 1994] T. Baudel, A Mark-based interaction paradigm for free-hand drawing. *ACM UIST '94 Symposium on User Interface Software and Technology*, ACM Press, 1994, 185-192.
- [Bewley *et al.*, 1983] William L. Bewley, Teresa L. Roberts, David Schroit, and William L. Verplank, Human factors testing in the design of Xerox's 8010 "Star" office workstation. *Human Factors in Computing Systems, CHI '83 Conference Proceedings*, 1983, 72-77.
- [Bier *et al.*, 1993] Eric A. Bier, Maureen C. Stone, Ken Pier, William Buxton, and Tony D. DeRose, Toolglasses and magic lenses: the see-through interface. *SIGGRAPH '93 Conference Proceedings*, ACM Press, 1993, 73-80.

- [Bier *et al.*, 1994] Eric A. Bier, Maureen C. Stone, Ken Fishkin, William Buxton, and Thomas Baudel, A taxonomy of see-through tools. *Human Factors in Computing Systems, CHI '94 Conference Proceedings*, ACM Press, 1994, 358-364.
- [Bolt, 1980] Richard A. Bolt, Put-that-there. *SIGGRAPH '80 Conference Proceedings*, ACM Press, 1980, 262-270.
- [Bolt and Herranz, 1992] Richard A. Bolt and Edward Herranz, Two-handed gesture in multi-modal natural dialog. *ACM UIST '92 Symposium on User Interface Software and Technology*, ACM Press, 1992, 7-14.
- [Buxton, 1986] William Buxton, There's more to interaction than meets the eye: some issues in manual input. In: *User Centered System Design*, D. A. Norman and S. W. Draper (Eds.), Lawrence Erlbaum, 1986, 319-337.
- [Buxton and Myers, 1986] William Buxton and Brad A. Myers, A study in two-handed input. *Human Factors in Computing Systems, CHI '86 Conference Proceedings*, ACM Press, 1986, 321-326.
- [Buxton, 1990] A three-state model of graphical input. *Proceedings of INTERACT '90: The IFIP Conference on Human-Computer Interaction*, Elsevier Science Publishers (North-Holland), 1990, 449-456.
- [Charwat, 1992] H. J. Charwat, *Lexicon der Mensch-Maschine-Kommunikation*. Oldenbourg, 1992.
- [Chatty, 1994] Stéphane Chatty, Extending a graphical toolkit for two-handed interaction. *ACM UIST '94 Symposium on User Interface Software and Technology*, ACM Press, 1994, 195-204.
- [Christian and Avery, 1998] Andrew D. Christian and Brian L. Avery, Digital Smart Kiosk project. *Human Factors in Computing Systems, CHI '98 Conference Proceedings*, ACM Press, 1998, 155-162.
- [Cohen *et al.*, 1989] Philip R. Cohen, Mary Dalrymple, Douglas B. Moran, Fernando C. N. Pereira, Joseph W. Sullivan, Robert A. Gargan Jr, Jon L. Schlossberg, and Sherman W. Tyler, Synergistic use of direct manipulation and natural language. *Human Factors in Computing Systems, CHI '89 Conference Proceedings*, ACM Press, 1989, 227-233.
- [Cohen *et al.*, 1997] Philip R. Cohen, Michael Johnston, David McGee, Sharon Oviatt, Jay Pittman, Ira Smith, Liang Chen, and Josh Clow, QuickSet: multimodal interaction for distributed applications. *Proceedings of the Multimedia '97, Fifth ACM International Conference on Multimedia*, ACM Press, 1997, 31-40.

- [Coutaz, 1987] Joëlle Coutaz, PAC: An object-oriented model for dialog design. *Proceedings of INTERACT '87: The IFIP Conference on Human Computer Interaction*, 1987, 431-436.
- [Coutaz *et al.*, 1993] Joëlle Coutaz, Laurence Nigay, and Daniel Salber, The MSM framework: a design space for multi-sensori-motor systems. *Proceedings of EWHCI '93, Third East-West International Conference on Human-Computer Interaction, Lecture Notes in Computer Science*, 753, Springer-Verlag, 1993, 231-241.
- [van Dam, 1997] Andries van Dam, Post-WIMP user interfaces. *Communications of the ACM*, 40 (2), 1997, 63-67.
- [Dillon *et al.*, 1990] Richard F. Dillon, Jeff D. Edey and Jo W. Tombaugh, Measuring the true cost of command selection: techniques and results. *Human Factors in Computing Systems, CHI '90 Conference Proceedings*, ACM Press, 1990, 19-25.
- [Feldman and Rimé, 1991] Robert S. Feldman and Bernard Rimé (Eds.), *Fundamentals of Nonverbal Behavior*. Cambridge University Press, 1991.
- [Fitts, 1954] Paul M. Fitts, The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47 (6), 1954, 381-391.
- [Fitzmaurice *et al.*, 1995] George W. Fitzmaurice, Hiroshi Ishii and William Buxton, Bricks: Laying the foundations for graspable user interfaces. *Human Factors in Computing Systems, CHI '95 Conference Proceedings*, ACM Press, 1995, 442-449.
- [Fitzmaurice and Buxton, 1997] George W. Fitzmaurice and William Buxton, An empirical evaluation of graspable user interfaces: toward specialized, space-multiplexed input. *Human Factors in Computing Systems, CHI '97 Conference Proceedings*, ACM Press, 1997, 43-50.
- [Flowers, 1975] Kenneth Flowers, Handedness and controlled movement. *British Journal of Psychology*, 66, 1975, 39-52.
- [Foley *et al.*, 1990] James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes, *Computer Graphics: principles and practice*, 2nd edition. Addison-Wesley, 1990.
- [Fowler, 1992] B. Fowler, Geometric manipulation of tensor product surfaces. *Computer Graphics*, 26 (1), 1992, 101-108.
- [Gross and Do, 1996] M. D. Gross and E. Yi-Luen Do, Ambiguous intentions: a paper-like interface for creative design. *ACM UIST '96 Symposium on User Interface Software and Technology*, ACM Press, 1996, 183-192.

- [Guiard, 1987] Yves Guiard, Asymmetric division of labor in human skilled bimanual action: the kinematic chain as a model. *The Journal of Motor Behavior*, 19 (4), 1987, 486-517.
- [Heilig, 1955] Morton L. Heilig, El cine del futuro (The cinema of the future). *Espacios*, 23-24, Apartado Postal Num. 20449, Espacios SA, Mexico, 1955.
- [Hinckley *et al.*, 1997a] Ken Hinckley, Randy Pausch, Dennis Proffitt, James Patten, and Neal Kassell, Cooperative bimanual action. *Human Factors in Computing Systems, CHI '97 Conference Proceedings*, ACM Press, 1997, 27-34.
- [Hinckley *et al.*, 1997b] Ken Hinckley, Randy Pausch, and Dennis Proffitt, Attention and visual feedback: the bimanual frame of reference. *Proceedings of SI3D '97, 1997 ACM Symposium on Interactive 3D Graphics*, ACM Press, 1997, 121-126.
- [Hinckley *et al.*, 1998a] Ken Hinckley, Randy Pausch, Dennis Proffitt, and Neal F. Kassell, Two-handed virtual manipulation. *ACM Transactions on Computer-Human Interaction*, 5 (3), 1998, 260-302.
- [Hinckley *et al.*, 1998b] Ken Hinckley, Mary Czerwinski, and Mike Sinclair, Interaction and modeling techniques for desktop two-handed input. *ACM UIST '98 Symposium on User Interface Software and Technology*, ACM Press, 1998, 49-58.
- [Hinckley and Sinclair, 1999] Ken Hinckley and Mike Sinclair, Touch-sensing input devices. *Human Factors in Computing Systems, CHI '99 Conference Proceedings*, ACM Press, 1999, 223-230.
- [Hoffmann, 1991] Errol R. Hoffmann, A comparison of hand and foot movement times. *Ergonomics*, 34 (4), 1991, 397-406.
- [Ishii and Ullmer, 1997] Hiroshi Ishii and Brygg Ullmer, Tangible bits: towards seamless interfaces between people, bits and atoms. *Human Factors in Computing Systems, CHI '97 Conference Proceedings*, ACM Press, 1997, 234-241.
- [Jacob, 1993] Robert J. K. Jacob, Eye movement-based human-computer interaction techniques: toward non-command interfaces. In: *Advances in Human-Computer Interaction*, Vol. 4, H. R. Hartson and D. Hix (Eds.), Ablex Publishing, 1993, 151-190.
- [Jacob *et al.*, 1999] Robert J. K. Jacob, Leonidas Deligiannidis, and Stephen Morrison, A software model and specification language for Non-WIMP user interfaces. *ACM Transactions on Computer-Human Interaction*, 6 (1), 1999, 1-46.

- [Kabbash *et al.*, 1993] Paul Kabbash, I. Scott MacKenzie and William Buxton, Human performance using computer input devices in the preferred and non-preferred hands. *Human Factors in Computing Systems, INTERCHI '93 Conference Proceedings*, ACM Press, 1993, 474-481.
- [Kabbash *et al.*, 1994] Paul Kabbash, William Buxton and Abigail Sellen, Two-handed input in a compound task. *Human Factors in Computing Systems, CHI '94 Conference Proceedings*, ACM Press, 1994, 417-423.
- [Kandel and Schwartz, 1981] E. R. Kandel and J. R. Schwartz, *Principles of Neural Sciences*. Elsevier Science Publishers (North Holland), 1981.
- [Koons *et al.*, 1993] D. B. Koons, C. J. Sparrel, and K. R. Thorisson, Integrating simultaneous input from speech, gaze, and hand gestures. In: *Intelligent Multimedia Interfaces*, M. Maybury (Ed.), MIT Press, 1993, 257-276.
- [Krasner and Pope, 1988] G. E. Krasner and S. T. Pope, A description of the model-view-controller user interface paradigm in the Smalltalk-80 system. *Journal of Object-Oriented Programming*, 1 (3), 1988, 26-49.
- [Kurtenbach *et al.*, 1997] Gordon Kurtenbach, George Fitzmaurice, Thomas Baudel, and Bill Buxton, The design of a GUI paradigm based on tablets, two-hands and transparency. *Human Factors in Computing Systems, CHI '97 Conference Proceedings*, ACM Press, 1997, 35-42.
- [Krueger, 1983] Myron W. Krueger, *Artificial Reality*. Addison-Wesley, 1983.
- [Krueger *et al.*, 1985] Myron W. Krueger, Thomas Gionfriddo, and Katrin Hinrichsen, VIDEOPLACE – an artificial reality. *Human Factors in Computing Systems, CHI '85 Conference Proceedings*, ACM Press, 1985, 35-40.
- [Leganchuk *et al.*, 1998] Andrea Leganchuk, Shumin Zhai, and William Buxton, Manual and cognitive benefits of two-handed input: an experimental study. *ACM Transactions on Computer-Human Interaction*, 5 (4), 1998, 326-359.
- [Lipkie *et al.*, 1982] Daniel E. Lipkie, Steven R. Evans, John K. Newlin, and Robert L. Weissman, Star graphics: an object-oriented implementation. *Proceedings of SIGGRAPH '82, Computer Graphics*, 16 (3), 1982, 115-124.
- [MacDonald and McGurk, 1978] J. MacDonald and H. McGurk, Visual influences on speech perception process. *Perception and Psychophysics*, 24 (3), 1978, 253-257.
- [MacKenzie, 1992] I. Scott MacKenzie, Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, 7, 1992, 91-139.
- [Mackinlay *et al.*, 1991] Jock Mackinlay, Stuart Card, and George Robertson, A semantic analysis of the design space of input devices. *Human-Computer Interaction*, 5, 1991, 145-190.

- [Matsushita and Rekimoto, 1997] Nobuyuki Matsushita and Jun Rekimoto, HoloWall: designing a finger, hand, body, and object sensitive wall. *ACM UIST '97 Symposium on User Interface Software and Technology*, ACM Press, 1997, 209-210.
- [Maybury and Wahlster, 1998] Mark T. Maybury and Wolfgang Wahlster (Eds.), *Readings in Intelligent User Interfaces*. Morgan Kaufmann Publishers, 1998.
- [McGurk and MacDonald, 1976] H. McGurk and J. MacDonald, Hearing lips and seeing voices. *Nature*, 264, 1976, 746-748.
- [MIAMI, 1995] L. Schomaker, J. Nijtmans, A. Camurri, F. Lavagetto, P. Morasso, C. Benoît, T. Guiard-Marigny, B. Le Goff, J. Robert-Ribes, A. Adjoudani, I. Defée, S. Münch, K. Hartung, and J. Blauert, *A Taxonomy of Multimodal Interaction in the Human Information Processing System*. A Report of the Esprit Basic Research Action 8579 MIAMI. February, 1995.
<http://vonkje.cogsci.kun.nl/~miami/reports/reports.html>
 (checked 15 Nov 99)
- [Miller, 1997] Jim Miller, Intelligent software agents vs. user-controlled direct manipulation: a debate. *Human Factors in Computing Systems, CHI '97 Extended Abstracts*, ACM Press, 1997, 105-106.
- [Moran *et al.*, 1997] Douglas B. Moran, Adam J. Cheyer, Luc E. Julia, David L. Martin, and Sangkyu Park, Multimodal user interfaces in the Open Agent Architecture. *Proceedings of the 1997 International Conference on Intelligent User Interfaces (IUI '97)*, ACM Press, 1997, 61-68.
- [MultiGen, 1999] MultiGen Inc., SmartScene product home page.
<http://www.multigen.com/products/smarts.htm> (checked 15 Nov 99)
- [Myers, 1990] Brad A. Myers, A model for handling input. *ACM Transactions on Information Systems*, 8 (3), 1990, 289-320.
- [Neal *et al.*, 1989] J. G. Neal, C. Y. Thielman, Z. Dobes, S. M. Haller, and S. C. Shapiro, Natural language with integrated deictic and graphic gestures. *Proceedings of the 1989 DARPA Workshop on Speech and Natural Language*, Morgan Kaufmann, 1989, 410-423.
- [Neal and Shapiro, 1991] Jeannette G. Neal and Stuart C. Shapiro, Intelligent multi-media interface technology. In: *Intelligent User Interfaces*, J. W. Sullivan and S. W. Tyler (Eds.), ACM Press, 1991, 11-43.
- [Nigay and Coutaz, 1993] Laurence Nigay and Joëlle Coutaz, A design space for multimodal systems: concurrent processing and data fusion. *Human Factors in Computing Systems, INTERCHI '93 Conference Proceedings*, ACM Press, 1993, 172-178.

- [Nigay and Coutaz, 1995] Laurence Nigay and Joëlle Coutaz, A generic platform for addressing the multimodal challenge. *Human Factors in Computing Systems, CHI '95 Conference Proceedings*, ACM Press, 1995, 98-105.
- [Nigay *et al.*, 1993] Laurence Nigay, Joëlle Coutaz, and Daniel Salber, MATIS: a multimodal airline travel information system. SM/WP10, ESPRIT Basic Research Action 7040: AMODEUS, February 1993.
<http://www.mrc-cbu.cam.ac.uk/amodeus/sm.html> (checked 15 Nov 99)
- [Nielsen, 1993] Jakob Nielsen, Non-command user interfaces. *Communications of the ACM*, 36 (4), 1993, 83-99.
- [Norman, 1990] Donald A. Norman, Why interfaces don't work. In: *The Art of Human-Computer Interface Design*, B. Laurel (Ed.), Addison-Wesley, 1990, 209-219.
- [Olsen and Deng, 1996] Dan R. Olsen Jr. and Xinyu Deng, Inductive groups. *ACM UIST '96 Symposium on User Interface Software and Technology*, ACM Press, 1996, 193-199.
- [Oviatt, 1996] Sharon Oviatt, Multimodal interfaces for dynamic interactive maps. *Human Factors in Computing Systems, CHI '96 Conference Proceedings*, ACM Press, 1996, 95-102.
- [Oviatt, 1997] Sharon Oviatt, Multimodal interactive maps: designing for human performance. *Human-Computer Interaction*, 12, 1997, 93-129.
- [Oviatt *et al.*, 1997] Sharon Oviatt, Antonella DeAngeli, and Karen Kuhn, Integration and synchronization of input modes during multimodal human-computer interaction. *Human Factors in Computing Systems, CHI '97 Conference Proceedings*, ACM Press, 1997, 415-422.
- [Oviatt, 1999a] Sharon Oviatt, Mutual disambiguation of recognition errors in a multimodal architecture. *Human Factors in Computing Systems, CHI '99 Conference Proceedings*, ACM Press, 1999, 576-583.
- [Oviatt, 1999b] Sharon Oviatt, Ten myths of multimodal interaction. *Communications of the ACM*, 42 (11), 74-81.
- [Parke and Waters, 1996] Frederick I. Parke and Keith Waters, *Computer Facial Animation*. A K Peters, 1996.
- [Perkins *et al.*, 1997] Roderick Perkins, Dan Smith Keller, and Frank Ludolph, Inventing the Lisa user interface. *Interactions*, 4 (1), 1997, 40-53.
- [Pimentel and Teixeira, 1995] K. Pimentel and K. Teixeira, *Virtual Reality: Through the New Looking Glass*, 2nd edition. Intel/McGraw-Hill, 1995.

- [Qin and Terzopoulos, 1995] H. Qin and D. Terzopoulos, Dynamic manipulation of triangular B-splines. *Proceedings of the Third Symposium on Solid Modeling and Applications (SMA '95)*, ACM Press, 1995, 351-360.
- [Raisamo and R  ih  , 1996] Roope Raisamo and Kari-Jouko R  ih  , A new direct manipulation technique for aligning objects in drawing programs. *ACM UIST '96 Symposium on User Interface Software and Technology*, ACM Press, 1996, 157-164.
- [Raisamo, 1998] Roope Raisamo, A multimodal user interface for public information kiosks. *Proceedings of 1998 Workshop on Perceptual User Interfaces (PUI '98)*, Matthew Turk (Ed.), San Francisco, November 1998, 7-12.
<http://research.microsoft.com/PUIWorkshop/Papers/Raisamo.pdf>
 (checked 15 Nov 99)
- [Raisamo, 1999a] Roope Raisamo, An alternative way of drawing. *Human Factors in Computing Systems, CHI '99 Conference Proceedings*, ACM Press, 1999, 175-182.
- [Raisamo, 1999b] Roope Raisamo, An empirical study on how to make direct manipulation more direct in drawing programs. *OzCHI '99 Conference Proceedings, 1999 Conference of the Computer Human Interaction Special Interest Group of the Ergonomics Society of Australia*, School of Information Studies, Charles Sturt University, 1999, 71-77.
- [Raisamo, 1999c] Roope Raisamo, Evaluating different touch-based interaction techniques in a public information kiosk. Report A-1999-11, Department of Computer Science, University of Tampere, 1999.
<ftp://ftp.cs.uta.fi/pub/reports/pdf/A-1999-11.pdf>
 (checked 15 Nov 99)
- [Raisamo, 1999d] Roope Raisamo, Evaluating touch-based area selection techniques in a public information kiosk. *OzCHI '99 Conference Proceedings, 1999 Conference of the Computer Human Interaction Special Interest Group of the Ergonomics Society of Australia*, School of Information Studies, Charles Sturt University, 1999, 169-171.
- [Raisamo and R  ih  , 1999] Roope Raisamo and Kari-Jouko R  ih  , Design and evaluation of the alignment stick. *Interacting with computers* (accepted for publication).
- [Raisamo et al., 1999] Roope Raisamo, Tero Kukola, Jouni Huovinen, Kalle Malin, and Poika Isokoski, New interaction techniques home pages.
<http://www.cs.uta.fi/research/hci/interact/> (checked 15 Nov 99)

- [Robbins *et al.*, 1999] Jason Robbins, Michael Kantor, and David Redmiles, Sweeping away disorder with the broom alignment tool. *Human Factors in Computing Systems, CHI '99 Extended Abstracts*, ACM Press, 1999, 250-251.
- [Salomon, 1990] Gitta B. Salomon, Designing casual-use hypertext: the CHI '89 InfoBooth. *Human Factors in Computing Systems, CHI '90 Conference Proceedings*, ACM Press, 1990, 451-458.
- [Sams *et al.*, 1998] M. Sams, P. Manninen, V. Surakka, P. Helin, and R. Kättö, Effects of word meaning and sentence context on the integration of audiovisual speech. *Speech communication* (in press).
- [Schmidt *et al.*, 1979] Richard A. Schmidt, Howard Zelaznik, Brian Hawkins, James S. Frank, and John T. Quinn Jr., Motor-output variability: a theory for the accuracy of rapid motor acts. *Psychological Review*, 86 (5), 1979, 415-451.
- [Shannon and Weaver, 1949] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, 1949.
- [Shepherd, 1988] G. M. Shepherd, *Neurobiology*, 2nd edition. Oxford University Press, 1988.
- [Shneiderman, 1982] Ben Shneiderman, The future of interactive systems and the emergence of direct manipulation. *Behaviour and Information Technology* 1 (3), 1982, 237-256.
- [Shneiderman, 1983] Ben Shneiderman, Direct manipulation: a step beyond programming languages. *IEEE Computer*, 16 (8), 57-69.
- [Shneiderman and Maes, 1997] Ben Shneiderman and Pattie Maes, Direct manipulation vs. interface agents. *Interactions*, 4 (6), 1997, 42-61.
- [Silbernagel, 1979] D. Silbernagel, *Tatchenatlas der Physiologie*. Thieme, 1979.
- [Sutherland, 1963] Ivan E. Sutherland, SketchPad: A man-machine graphical communication system. *AFIPS Conference Proceedings* 23, 1963, 323-328.
- [Taylor, 1989] R. M. Taylor, Integrating voice, visual and manual transactions: some practical issues from aircrew station design. *The Structure of Multimodal Dialogue*, M. M. Taylor, F. Néel and D. G. Bouwhuis (Eds.), Elsevier Science Publishers (North-Holland), 1989, 259-265.
- [Terzopoulos *et al.*, 1987] D. Terzopoulos, J. Platt, A. Barr, and K. Fleischer, Elastically deformable models. *Computer Graphics*, 21 (4), 1987, 205-214.
- [Thorisson *et al.*, 1992] Kristinn R. Thorisson, David B. Koons, and Richard A. Bolt, Multi-modal natural dialogue. *Human Factors in Computing Systems, CHI '92 Conference Proceedings*, ACM Press, 1992, 653-654.

- [Todor and Doane, 1978] John I. Todor and Thomas Doane, Handedness and hemispheric asymmetry in the control of movements. *Journal of Motor Behavior*, 10 (4), 1978, 295-300.
- [Ullmer and Ishii, 1997] Brygg Ullmer and Hiroshi Ishii, The metaDESK: models and prototypes for tangible user interfaces. *ACM UIST '97 Symposium on User Interface Software and Technology*, ACM Press, 1997, 223-232.
- [USB, 1999] Universal serial bus implementers forum home page.
<http://www.usb.org/> (checked 15 Nov 99)
- [Wacom, 1999] Wacom Technology Co., PenTools 3.0 Plugins home page.
<http://www.wacom.com/pentools/> (checked 15 Nov 99)
- [Wahlster, 1991] Wolfgang Wahlster, User and discourse models for multimodal communication. In: *Intelligent User Interfaces*, J. W. Sullivan and S. W. Tyler (Eds.), ACM Press, 1991, 45-67.
- [Waters and Levergood, 1993] Keith Waters and Thomas M. Levergood, DECface: an automatic lip-synchronization algorithm for synthetic faces. Technical report 93/4, Digital Cambridge Research Lab, September 1993.
<ftp://crl.dec.com/pub/DEC/CRL/tech-reports/93.4.ps.Z>
 (checked 15 Nov 99)
- [Weimer and Ganapathy, 1989] David Weimer and S. K. Ganapathy, A synthetic virtual environment with hand gesturing and voice input. *Human Factors in Computing Systems, CHI '89 Conference Proceedings*, ACM Press, 1989, 235-240.
- [Welford, 1968] A. Welford, *Fundamentals of Skill*. Methuen, London, 1968.
- [Wickens, 1992] C. D. Wickens, *Engineering Psychology and Human Performance*. Harper Collins, New York, 1992.
- [Yang *et al.*, 1998] Jie Yang, Rainer Stiefelhagen, Uwe Meier, and Alex Waibel, Visual tracking for multimodal human computer interaction. *Human Factors in Computing Systems, CHI '98 Conference Proceedings*, ACM Press, 1998, 140-147.
- [Zhai *et al.*, 1997] Shumin Zhai, Barton A. Smith, and Ted Selker, Improving browsing performance: a study of four input devices for scrolling and pointing tasks. *Proceedings of INTERACT '97: The IFIP Conference on Human-Computer Interaction*, 1997, 286-292.
- [Zimmerman *et al.*, 1995] Thomas G. Zimmerman, Joshua R. Smith, Joseph A. Paradiso, David Allport, and Neil Gershenfeld, Applying electric field sensing to human-computer interfaces. *Human Factors in Computing Systems, CHI '95 Conference Proceedings*, ACM Press, 1995, 280-287.

PAPER I

Roope Raisamo and Kari-Jouko Räihä,

A new direct manipulation technique for aligning objects in drawing programs. *ACM UIST '96 Symposium on User Interface Software and Technology*, ACM Press, 1996, 157-164. [Raisamo and Räihä, 1996]

Available online at:

[http://www.acm.org/pubs/articles/proceedings/
uist/237091/p157-raisamo/p157-raisamo.pdf](http://www.acm.org/pubs/articles/proceedings/uist/237091/p157-raisamo/p157-raisamo.pdf)

(checked 15 Nov 99 – requires ACM Digital Library subscription)

PAPER II

Roope Raisamo and Kari-Jouko Räihä,

Design and evaluation of the alignment stick. *Interacting with computers* (accepted for publication). [Raisamo and Räihä, 1999]

PAPER III

Roope Raisamo,

An alternative way of drawing. *Human Factors in Computing Systems, CHI '99 Conference Proceedings*, ACM Press, 1999, 175-182. [Raisamo, 1999a]

Available online at:

[http://www.acm.org/pubs/articles/proceedings/
chi/302979/p175-raisamo/p175-raisamo.pdf](http://www.acm.org/pubs/articles/proceedings/chi/302979/p175-raisamo/p175-raisamo.pdf)

(checked 15 Nov 99 – requires ACM Digital Library subscription)

PAPER IV

Roope Raisamo,

An empirical study of how to use input devices to control tools in drawing programs. *OzCHI '99 Conference Proceedings, 1999 Conference of the Computer Human Interaction Special Interest Group of the Ergonomics Society of Australia*, School of Information Studies, Charles Sturt University, 1999, 71-77. [Raisamo, 1999b]

Available online at:

<http://www.csu.edu.au/OZCHI99/>

(checked 15 Nov 99)

PAPER V

Roope Raisamo,

A multimodal user interface for public information kiosks.
Proceedings of 1998 Workshop on Perceptual User Interfaces (PUI '98),
Matthew Turk (Ed.), San Francisco, November 1998, 7-12. [Raisamo,
1998]

Available online at:

<http://research.microsoft.com/PUIWorkshop/Papers/Raisamo.pdf>

(checked 15 Nov 99)

PAPER VI

Roope Raisamo,

Evaluating different touch-based interaction techniques in a public information kiosk. Report A-1999-11, Department of Computer Science, University of Tampere, 1999. [Raisamo, 1999c]

Available online at:

<ftp://ftp.cs.uta.fi/pub/reports/pdf/A-1999-11.pdf>

(checked 15 Nov 99)

Published as a short paper in *OzCHI '99 Conference Proceedings, 1999 Conference of the Computer Human Interaction Special Interest Group of the Ergonomics Society of Australia*, School of Information Studies, Charles Sturt University, 1999, 169-171. [Raisamo, 1999d].

Available online at:

<http://www.csu.edu.au/OZCHI99/>

(checked 15 Nov 99)