Interaction Design Institute Ivrea

THESIS REPORT
May 22, 2006

# Hands-On

Shawn Bonkowski

s.bonkowski@interaction-ivrea.it
people.interactino-ivrea.it/s.bonkowski/thesis/

Primary Advisor:
Neil Churcher

Secondary Advisor:
Massimo Banzi

Report Advisor:
Simona Maschi

Directors and chair of the Examiners:
Neil Churcher and Heather Martin

# Hands-On

Shawn Bonkowski

## Abstract

This thesis is an investigation of how hand gestures can be used as a source of input in an indirect computing environment. The premise behind this thesis is that hands can also provide multiple dimensions and multiple levels of granularity to graphical environments that allow for a more direct and more immediate relationship with the computing system.

# Contents

# 1 Introduction

The main goal of this thesis has become an exploration of ways of using the rich and broad range of hand gestures in an indirect computing environment. The distinction between direct and indirect interfaces is simple. With a direct interface, the object does as it appears, a hammer is used to hammer. With an indirect interface, there is something to mediate our intentions. The text editor window on your computer provides you the facilities of a typewriter. This thesis initially began as a loose set of ideals surrounding the idea that humans, through gestures, hand waving or body language, can communicate a great deal of information. Other people can often understand with little explanation the meaning of a gesture based on the context of that gesture. A police officer standing in a traffic intersection can control the flow of traffic with a limited amount of body language.

As stated above, hand gestures have become the main interest of this thesis. About halfway through the thesis exploration the idea of hand gesture controls became the major interest of this thesis. Initially, it was confined within a specific application context. However, that context limited the exploration of hand controls through the requirements of the application at large. A number of interface items were proposed within that context, but once the application context was dropped, a specific focus of multi-granular and multi-dimensional control interfaces emerged. With these new potentials came a number of new ideas and possibilities emerged to extend the way we interact with computing systems.

Another motivation behind the *Hands-On* concept, is the mediation that is a result of the input systems we have available in indirect computing systems. There are a number things we do in the analogue world that we have recreated digitally but in the ways that we are required to interact with these digital versions, we have lost some conveniences or comforts. It is through the extension of the keyboard and mouse paradigm that we can possibly reclaim the subtleties lost.

The theme of multi-granular and mutli-dimensional input is present in all the concepts listed in chapters 2 through 6. Some of these concepts are more practical or acceivable than others but in the end they all rely on a large dynamic range of input intentions that can be expressed with some form or subset of body language.

Oren Horev's thesis, tentatively titled *Shape Shifters*, is concerned with communicating to the hand, while this thesis is about communications from the hand in the form of gestures, actions or movements.

## 1.1 Purpose

The purpose of this report is to summarize the activities and evolution involved in the development of this thesis. It is to serve as more of a companion to the thesis project rather than the thesis itself. It contains the initial ideas and concepts, and if implemented, the details of the implementations.

## 1.2 Document Overview

Section 1.3 *Early Thoughts* is a collection of early notions of what this thesis should be about or some of the values that should be contained in the final outcome of this project. These ideas were distilled or fleshed out in the concepts presented in chapters 2 through 6.

Chapters 2 through 6 each contain a different concept that was explored even if only briefly. In each of these chapters there is a major subsection called *Concept Generation, Design and Implementation*, is a description of the different concepts. The other major subsection of each chapter is *Technological and Economic Study*. This subsection ontains initial practical considerations for the implementation of the concepts.

Chapters 7 and 8, *Final Analysis and Evaluation* and *Conclusion* are presently are largely just placeholders and will be completed in later versions of this thesis report. These sections do however, contain a re-examintion of the this thesis and the questions it hopes answer once completed.

The *Sources* chapter contains a list of sources that influenced the evolution of this thesis.

### 1.2.1 Concept Generation, Design and Implementation Overview

These subsections in each chapter provide an introduction into one of several concepts that were considered and developed for this thesis. Not all of the concepts were taken forward, in fact only one will be developed to completion. Additionally, only some of the concepts and directions arising from the *Hands-on Desktop* will be the main focus of this thesis. The other concepts listed here are other explorations into what could be done when body language or gestures are considered as the main input source of a system.

The general idea behind all of these concepts presented in this section was the use of human gestures or body language as an input mechanism. The first concept *Hands-on Desktop* is the only indirect interface concept, while the other concepts are all direct interface concepts. The terms indirect and direct should not be confused as pertaining to the immediacy of interaction with a system, rather they describe a specific quality of the interaction. Using a mouse or keyboard on your keyboard causes actions to occur within the application you are using but these actions are mediated through the computer’s display and windowing system. Thus the keyboard and mouse are examples of indirect interfaces. The flip cover on a mobile phone is a direct interface for turning on the main features of the phone. There is a one-to-one, or direct, relationship between the interface and the result of action of the interface. The concepts, while some have not been developed in depth, explore some of the possibilities and potentials of both types of interfaces.

The specific aspect that has become the main focus of this thesis is outlined in chapter 2 *Hands-On* and more. The reason for this placement in this report is that it follows the development of the *Hands-on Desktop* concept and is a single point within the larger concept. The name has not changed because it is a more focused examination of the most significant aspect of the overall concept outlined in section 2.1 and is the portion of the concept that is most literally related to the name *Hands-on Desktop*.

### 1.2.2 Technological and Economic Study Overivew

Each concept chapter in this report contains a *Technological and Economic Study* section. All of concepts mentioned in chapters 2 through 6 of this report could use the same technology for its input system. All the concepts rely on some sort of gesture recognition system for decoding the users movements and doing something meaningful with those actions. The Eyesweb platform, developed at the University of Genoa, Italy, is a powerful and flexible gesture interpretation system. One of Eyesweb's core strengths is a video processing system that allows for real time processing of video streams. An additional benefit of the Eyesweb system is its ability to use just about any video source as an input, ranging from a €20 USB webcam to professional digital camera solutions costing thousands. This hardware independence allows for inexpensive prototyping to prove that concepts can work without having to spend large sums of money to do so.

The Eyesweb platform currently only runs on Microsoft Windows based machines as it relies heavily on Intel's Open Computer Vision (OCV) libraries. As Apple machines transition their processors from the PowerPC architecture to the Intel Pentium architecture it should be possible to use the OCV libraries eventually on a Mac. The Eyesweb platform, as of version 4, is moving to a less Microsoft Windows dependant implementation and thus should become easier to port to other systems. While Eyesweb system development is not of direct impact on this thesis, the impacts of the systems that it can run on will have an impact on what hardware any of the concepts mentioned in section 2 could be implemented.

## 1.3 Early Thoughts

I've been using different technologies for far too long now, and while I find they provide a useful tool set to accomplish myriad tasks, there is something lacking in my experiences. With different technologies, we are necessarily bound to their interfaces in order to use them. What I am interested in is discovering if we can move away from their specific interfaces and have the interfaces come to us, the users. Since most interfaces work well enough with respect to what they do, this is not an attempt to redefine the way we use things but rather an exploration into augmenting the way we use things.

What I'm looking for is a means to improve the way we do some "analogue" things digitally. I think that with some extra intelligence in objects or environments, we can immerse ourselves more completely in what we are doing rather than being interrupted by how we are doing it.

The motivation behind exploring gestural interfaces is also a matter of instinct. We have learned to turn a light on and off using the light's switch. It wouldn't be natural to use a keyboard and a string of commands to work the light switch. However, we might gesture to someone to turn on a light by pointing at it. Why can't the light recognise this gesture and turn itself on. Or maybe an aware environment recognises this gesture and turns it on for us.

One of the most immersive technology I've experienced is while playing Minesweeper on my computer. Muscle memory takes over the onscreen pointer movement through unconscious mouse movements. Once you've played the game enough to recognise number patterns and their implications (ie. no longer doing any analysis or reasoning on the pattern of numbers on the game board), the game was reduced to spotting patterns, marking mines, and finishing the game. If I chose the wrong box and lost the game because of a reasoning shortfall or mistake this was no big deal. Dirt in the mouse mechanism had a significant impact because it interfered with and broke the connection between my thinking and the cursor position onscreen. I have yet to find out why this would cause the sensation of a physical impediment in my arm. Playing Minesweeper in Windows XP running on a virutal machine on my iBook made my arm "feel" like it was throwing cinderblocks underwater - very heavy and very sluggish and possibly the least ideal environment to play the game.

How does this immersion or instictive-ness relate to the world?

The fascination some have with analogue Post-It notes in a digital age is to me if not paradoxical, than at least inefficient. The wonderful thing about Post-Its is they are very quick and flexible. Grab one, stick it in some sort of cluster of related ideas to form a group. Pick it off the wall, stick it in another group or even the garbage if it no longer works at all. My problem with this easy to use system is that its very difficult to take the wall away with me. Or even archiving the wall. There exists software like Microsoft Visio or Omnigroup's Omnigraffle that can group and cluster like the Post-It but trapped inside a small screen. (Both of the these applications do go far beyond wall based brainstorming.) Another partial solution are digital white board systems. These systems can capture all the data (if you write on the Post-Its (within the limits of the digital system) but don't provide any of the nice features that the aforementioned software systems can provide. What I would like is the immediacy of writing on the wall and additional functionality of the software provides while being able to use any wall to do this (and of course, easily take the content on the wall with me)

Instinctive and immersive interfaces are not simply gestural input systems. Both the input to the system and the system itself need to be tuned for this style of interaction. Without appropriate feedback, and the feedback systems may not yet really exists - I don't know yet of any walls that can behave like computer screens.

One means of communication that people have amongst each other is through a body gesture language. These gestures can be as simple as the shaking or nodding of your head to a complete sign language to replace spoken communication. While these gesture languages differ in complexity, effective communication can still be accomplished. One question is can different technologies become aware of human gestures and respond accordingly? Can a ubiquitous (even if localised) gesture language be used across unrelated devices?

Gesture input systems are not new. Different groups at Sony's Computer Science Labratory (CSL) have conducted explorations using gesture systems in media and environemtn manipulation. Recently, defence contractor Raytheon has been

working on implementing the gesture data manipulation system from the Future Crimes Unit
from the film "The Minority Report."

I don't yet know what my explorations in immersive and instinctive interfaces might yield. One possibility is a video input system that can track people in their environments to do the little things for them based on the flick of a wrist or directed point of the finger.

During an internship at Digit London, I worked on a multi-touch display system. The system had a large rear-projected surface at and a camera system that monitored light levels on the rear of the projection screen. Basically the camera served as a touch input system for what was being projected on the same screen.

# 2 Hands-On

Hands-On became the primary focus of this thesis. The indirect computing environment has been designed around the technology at the time of its creating for the tasks that were common at the time, such as word processing and spreadsheets. While that model serves well for a lot of applications that are still used today, the computing uses and needs have evolved significantly since the first computer mouse and graphical user interface were created. Today we have far more rich media content and graphical representations of information than we have ever had before, but we are still largely treating it with the same input paradigm of the keyboard and mouse. This provides and interesting area for exploration and has lead to an examination of what two hands might be able to do in an indirect computing environment.

Hands-On, started as a table computing idea but has since taken a step back from a specific application to look at the possibilities that using our hands can afford computing and how the keyboard and mouse paradigm can be extended. While table sized environments provide a larger dimension of scale, it is still the gestures or the actions that can be performed that is of interest, regardless of whether the system is coffee table sized or tablet computer sized.

In section 2.1, the exploration or evolution of the idea of hand gestures as input is presented in almost the order it took place, including earlier notions of an overly complex system that overshadowed the idea of hands as input devices. However, the motivations and ideas behind the complex system provided the roots from which the Hands-On concept grew. Section 2.2 outlines some of the technical and economic concerns of this type of interface. Some of the ideas presented in this chapter may be considered future casting or blue sky concepts, so the technological and economic impact is part fact based on current technologies and part speculation of how far reaching these concepts could spread.

## 2.1 Concept Generation, Design, Implementation

*Hands-on Desktop* (HoD) has undergone the most evolution of any of the ideas presented in this chapter. It started as a collaborative music and sound manipulation loosely inspired by an early idea in Chia-Ying Lee's thesis process. Since then it evolved into a more general purpose editing and visual media manipulation system. With either of these incarnations, the idea was to look at how you might manipulate or modify either the sound or the image with your hands.

A larger encompassing concept is presented in Section 2.1.1 while the detailed exploration begins in Section 2.1.2. and continues in Section 2.1.3 with a specific look at the possibilities of a multi-dimensional and multi-granular input system.

### 2.1.1 The table projects

The evolution of this concept is worth some mention as the issues, sometimes just pet-peeves, have shaped the desire this thesis and its direction towards two handed input systems.

The sound table grew partially out of frustration with a number of digital forms of analogue tools. One frustration was the inability to use more than one slider at a time the way you would if you were using a real mixing board. Watching DJs doing their thing, it is immediately obvious that they would not be able to accomplish the same thing if they were using a screen interface. One mouse can possibly provide the same non-linear action that you might have with a real slider on a mixing board or while scratching a record on a turntable but as soon as you try to do more than one thing at a time, you need another input device. Keyboards have many possible input triggers but do not have any of the fine grain or variable qualities of moving a mouse. A number of solutions have been created to address this problem. A lot of them look like analogue mixing boards but they are just a collection of manual sliders with a USB interface to the host machine. So the metaphor has jumped onto the screen and off again into a box mimicking the original analogue object. The new physical form then lacks some of the flexibility a digital system could afford it. And the cycle continues.

Figure 2.1 Music Manipulation Table Early Sketch

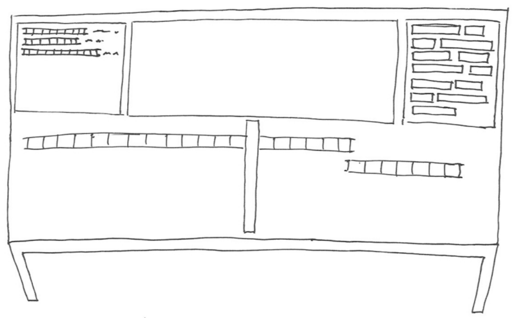

Figure 2.2 Editing Table Early Sketch

Most of my personal observations of the limitations of using digital audio systems came about while editing video and short films. The trend towards digital non-linear editing tools has been a welcome one. However, the attention and focus required with linear editing systems has been lost with the conveniences of the new tools. The idea behind evolving the table into video editing or manipulation station came out of the desire to have a better and more direct relationship with the material in the system. With non-linear systems it is too easy to make disposable decisions and so there is a tendency to become a lazy about know your material. If its not right you can fix it later. As opposed to knowing your material and getting it closer to right the first time. So the idea was to create an editing table that laid out the video material in front of you in a way that would make it more visually accessible and provide the experience of more directly manipulating the material.

### 2.1.2 Hands-on the interface

Section 2.1.1 outlined what the concept was about initially. What turns out to be a more interesting exploration is the relationship between what your hands are doing and what you would like the system to be doing in response to that. While the current Windows, Icons, Menus, and Pointer (WIMP) model works really well with a traditional keyboard and mouse system it doesn not take advantage of a person's whole hand. Generally, people have 10 fingers, not one, and while a mouse travels vertically (x-axis) and horizontally (y-axis) across the screen, hands do at least that much when driving a mouse, but can also travel up and down off the table which would be equivalent to travelling into or out of the screen (z-axis). Additionally, the orientation of the hand may also provide some information that is lost when using a mouse. The completely round mouse that Apple released with the first generation iMac demonstrated what can go wrong when rotation is added to a computer mouse – the cursor onscreen just did not go where you wanted it to.

The questions became those about how exactly would you use your hands to manipulate the material or the controls. With strong and weak hands potentially doing different things what are all the combinations and possibilities of interaction with the system. Bill Buxton's research at Alias Wavefront in interface design focused on the philosophy that the artist should not have to leave their work to go to a menu to do something. So following that philosophy what controls can be created for the strong and weak hands that work naturally and efficient for the user so that they can get on doing what they are doing. Other problems arise from the strong and weak hands – mainly is the user right or left handed? As a result should the menus and their organisation change to reflect this? With a standard mouse you can remap the buttons so the click function is under the index-finger of the strong hand. In practice however, I've witnessed lefties adjust their grip on a mouse so their index finger lands on the left button anyway. Apple computers are known for their single mouse buttons – perhaps to avoid the issue altogether.

Assuming that interactive table systems will become more prevalent, especially if the rise table computing increases the same way that monitor sizes have increased then some of the traditional boundaries in our interactions will computing systems will change. No longer will our input devices, mice and keyboards for example, be below our line of sight, while the results of our labours directly in our light of sight. If we begin to work on our output device, ie hands-on the desktop, then some new questions emerge. Will we still require tactile feedback from our input devices or will the visual feedback be sufficient? Will using our hands alone be satisfying or will we still want a physical input device. How will the relationship between strong and week hands change with respect to using the computing system? These questions are the basis for what this thesis hopes to determine. To validate the results of physical token versus hands only, a number of input and usage tests will be carried out to compare the different combinations of hands and devices within a narrow usage context.

At present some user research has been conducted into the actions and gestures used by different people to accomplish different tasks. Early control and menu structure layout has also been sketched out. These sketches are currently undergoing development and iterative testing to improve on the design and layout of controls.

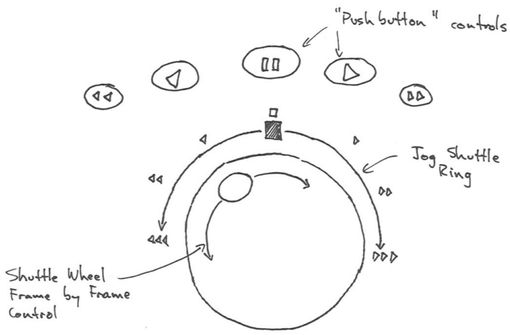


Figure 2.3 Jog Shuttle control sketches

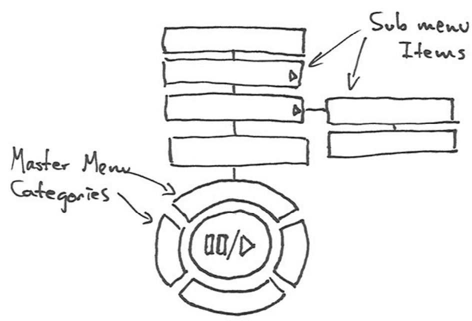


Figure 2.4 Early Master Menu control sketch

## 2.1.3 Hands-On – Multi-dimensional and Multi-granular

If we take a close look at the things that keyboard and mice were intended to serve and the time in which they were implemented we can see how interfaces were shaped. Initially computers were scientific machines, spreadsheet appliances, word processors, and other largely text based applications. Since then our computing environments have become rich with images, media, click boxes, drop down lists, and other interface objects that are tailored to a single click point afforded by the computer mouse. What if, for a moment we forgot about the mouse and looked at what our hands could do instead.

One of the more interesting and useful things about your hands is that they can do more than one thing at the same time. The can rotate, fingers can expand or contract, and they move about all at the same time. It is not difficult to see how a hand can replace a computer mouse: the position of the hand on the system replaces the mouse position, and a finger tapping the surface replaces a mouse click. However, that is the extent of the similarities in action. And the hand can do more still.

A simple example of hands doing more than one thing at a time is easier to demonstrate not with one hand but with two hands working at the same time. It is analogous to using a system that does not have one mouse but has two.

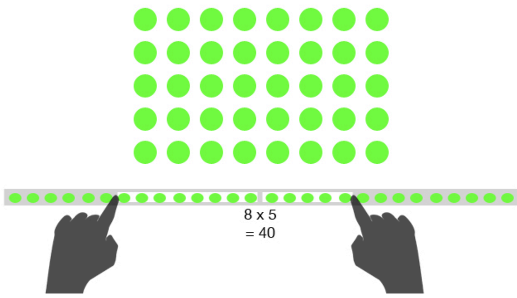


Figure 2.5 Hands-On Math 8*5=40

In figure 2.5, there are two hands each controlling the value of an operand in a mathematical equation. In this case the left hand has the value 8 and the right hand the value 5. But each of these values can be manipulated separately and independently of one another at the same time. This simple example highlights that very simply we can control multiple dimensions of a system at the same time.

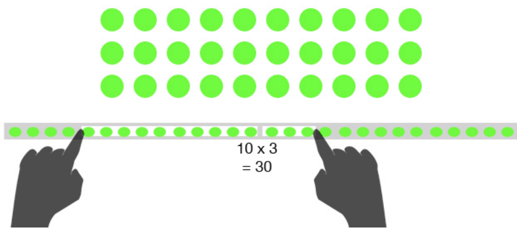


Figure 2.6 Hands-On Math 10*3=30

In Figure 2.6, purely by coincidence, both hands have shifted two values to the left to 10 and 3. Expanding this example with different mathematical operations we can quickly see what the effects of modifying one operand are. Compare simple hand movements to typing out the equation for evaluation a second time on a calculator, or clicking to select the value, change the value, updating the result to reflect the first change, and then repeating these steps to update the other side of the equation. Sliding your hands back and forth on the interface is immediate, while the keyboard and mouse method requires several distinct actions away from the task.

We can take this immediate action and extend it even further. In the Graphical Object Manipulation prototype (not shown) the colour picker is implemented as a tool for the left hand. The colour hue is mapped to the rotation of the left hand. The intensity of that colour is mapped to the finger extension of the left hand. So as the left hand rotates and fingers are extended or contracted the colour changes both in hue and in intensity. At the same time the right hand, is mapped to 3 different controls. The right hand mapping is very simple, position of hand to object position, object scale to expansion and contraction of the fingers, and the relative rotation of the hand to the relative rotation of the object. In this way, there are 5 aspects of an object that are directly and immediately modifiable. If we compare this to the traditional mouse and keyboard interfaces to perform the same action, there are numerous dialogs, menus, and individual values that need to be changed to accomplish the same modifications to a graphical object. There is an added bonus of being able to work on the object itself as opposed to colour swatches and other interfaces that remove the user from the task they are involved with.

In concert with well designed and ordered context menus, the ability to choose tools that quickly map to multiple dimensions, each with their own granular possibilities, these multi-demsional gestures can provide a new mechanisms to interact with graphical objects in the interface.

## 2.2 Technological and Economic Study

The display technology in many interactive tables is often one of the most expensive components of the system. While video projectors have come down in price, they typically still cost as much or more than the average desktop computer. Their cost will eventually come down but they will still remain a significant portion of the cost of an interactive table.

One of the most promising technologies to emerge that could take the interactive table potentials off the table is the Wedge display from Cambrige Flat Panels (UK). Using a specially shaped piece of glass the orientation of the table can be moved to wherever you want, without the restrictions and limitations of the space required by a projector to project the image. With the Wedge display the image is projected through the display itself. In addition to being a displace surface, the Wedge display can also act medium through which the camera input system can gather the images necessary for processing hand gestures. Basically it's a fancy piece of glass and light can travel through glass in many different directions.

The technical and economic aspects of this concept are difficult to define at this time because the concept is undergoing continual development and re-design. It will employ an expensive display technology, a video projector. It will require a computer running the Eyesweb system to handle the video input system. It will require development of hand gesture specific recognition rules for Eyesweb. It will require a desk sized glass table.

As for the economic implications of this system, as the price and size of projectors is reduced, and if the wedge display price is reduced to within reason (currently a 40cm display has a cost of approximately €9000), then using hand gestures as an input system may enter into common if not exotic and exclusive usage. The video cameras and computers are the least of the cost to the wide spread use of this type of input system.

The other option that is currently not technically feasible is a multi-touch sensitive tablet computer. With this potential technology, the range of potential applications opens up significantly. It would possibly require the development of an input protocol standard to communicate the actions of the hand to remote systems. With such a protocol and tablet system in place, any computer user could possibly make use of such an input paradigm and any of the ideas (and their logical extensions) would become possible for the masses as it removes the need for furniture sized computing appliances.

With respect to software or system potentials, it could enable all of the 3d visualisations, virtual environments, and other tech-bubble promises of the late 1990s and early 2000s. Economically this would impact consumers as and systems developers while the technology was being adopted, but ideally this usage paradigm would become common place and a natural included extension of the keyboard and mouse attached to almost every personal computer.

# 3 Fix My Swing, Yoga Me

## 3.1 Concept Generation, Design, Implementation

*Fix My Swing* and *Yoga Me* are two very similar concepts having to do with physical activity training. The first *Fix My Swing* (FMS) was about using a visual feedback system to help a user improve their skills using golf as its example activity. The second *Yoga Me* (YM) was more about training someone with a new skill or activity, in this example the skill was yoga.

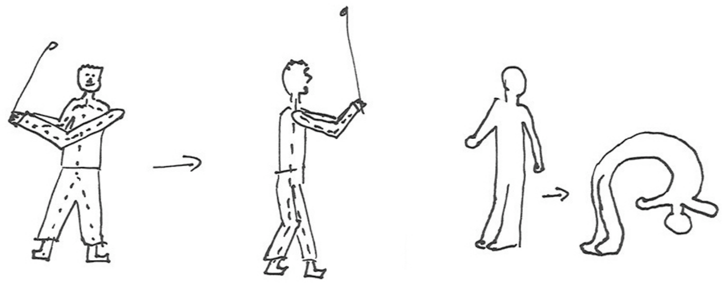

Figure 2.5 Early Fix My Swting and Yoga Me sketches

The basic idea behind FMS and YM was that the user could see an image of themselves in real time with an overlay of the correct movement. Both involved a video based gesture input system to monitor their moves and also to control the system. The video input in each system would then be interpreted by the system, and information relative to the activity would be overlaid onto the user's own image. The output information of course would need to have a convenient output display for the user of the system.

The value behind FMS was a system that could help the user self-correct their actions. Every golf television show seems to feature a segment on how to improve your swing – some trick or technique. With FMS you could load up a correct swing, or even one from your favorite player, and fix your swing through visualizations and analysis created by the system for you.

*Yoga Me* has a slightly different value. Rather than helping you correct something you are doing wrong or that needs improvement, YM is a system designed to teach you the basics of a new skill. First the system would demonstrate to you a movement and then you could compare your movement to the one demonstrated. Throughout the lesson your moves and actions would be continuously compared to the lesson movements and visual feedback provided so that the user could adjust and correct their moves in progress.

## 3.2 Technological and Econmic Study

The technology required for these concepts is fairly straightforward. A video input system, a computer to process the video input, and a display system for guiding the user in their actions.

Depending on the abilities to extract 3D information about the user's movements, the video input system could range from a single camera to a dual camera solution. Positioned correctly and if easily calibrated, then even multiple cameras would not be a problem for the end user. The economic implication of this is significant as it increase the chances that the end user would require no significant amounts of special purpose equipment to provide input to this system.

Due to available technologies, these concepts unfortunately have less than ideal display solutions. The visual feedback system providing that would provide you with the instruction or correction of your action would be one of three likely options: a computer monitor; a video projection; or computer display goggles. Each of these options has their own benefits and detractions.

Using a computer monitor would be the least expensive option for implementing the system but is perhaps the most restrictive and limiting for the user. The first restriction is the size of the screen that is small compared to the other display options. This alone may make it difficult for the user to see what movement or correction to the their movement they should be doing. The next restriction created by a monitor is that of position. If the movement or action requires significant changes in position, looking back to the screen will likely take the user of a correct position or prevent them from completing a movement correctly. While these limitations could be overlooked in if the user were to complete and then review their actions, this would defeat the real-time and immediate feedback benefits of the system.

The second option of a video projection would likely overcome some of the shortcomings of using a computer monitor. A larger display would make it easier for the user to see smaller details of their motions or actions and how they should be corrected. However, it does not overcome the issues of the positioning and the sightlines to the projection that create side effects such as having to turn around to see the screen. Video projections also represent the most costly display technology. So while it may be overall easier to see detail, the position and cost of video projections are still less than an ideal display solution.

The third potential display solution is computer display goggles. The only serious drawback is that cables leading to the goggles may tangle up the user of the system. This physical constraint may not be an issue if the movements or actions are relatively simple and do not the movement of the user. Until wireless VR goggles become available and reasonably affordable, display goggles will only be useful for limited user actions.

# 4 Remote Manipulation

## 4.1 Concept Generation, Design, Implementation

*Remote Manipulation* (RM) was a concept originally presented as *Shaky Hand Surgery* but has since been renamed to *Remote Manipulation.* With RM the idea was to use simple and inexpensive video systems to track a user's movements and mimic them on an expert system.

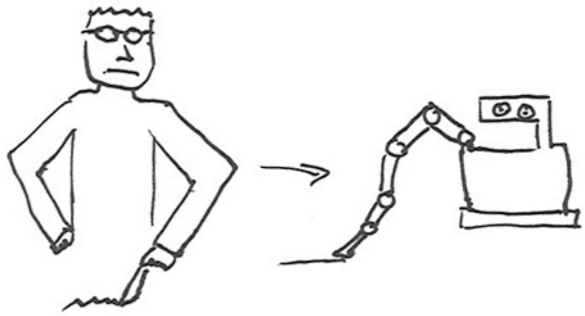

Figure 2.6 Early Shaky Hand Surgery concept sketch

The main idea behind remote manipulation was to contain as much expert knowledge in the machine performing the actual task. In the case of remote surgery, there is a pair of machines, one transmission machine and one receiving surgical machine. In this case the pair of machines is likely to be quite expensive as currently there is feedback sent to the transmission machine for the operator to 'feel'. However, if receiving machine had been programmed with the technical precision needed for an operation then there would be less need for feedback to the operator. Additionally, the system could then filter out incorrect movements or actions made by the operator of the system.

In order to accomplish this, the system transmitting the instructions to the receiving end of the system would need to understand the intent of the operator. This would require the sending half of the system to capture the gestures and then transmit the intent of those gestures to the receiving end of the system.

The operator of the system would require feedback of some sort in the system. A simple video feed could provide that response to the operator. Once, the range of movement is mapped, and there is no reason why movement scaling could no be calibrated, then the operator could perform a task from anywhere video conferencing can take place. The value in the context of remote surgery then is that you don't need a doctor at a particular facility, but that doctor becomes available anywhere in a time of need.

While surgery might be out the range of skills of the average person, the idea remains simple. We can, using our hands communicate a large number of simple actions that when sequenced together can form a complex set of actions. While not as complex as surgery think about all the actions that you perform in tying your shoes. Each portion of the actions required to tie a shoelace are not complex in them selves, a few crosses, a loop, feeding one string through another. This is, admittedly, a simplistic comparison but it remains that the end result of simple constituent movements can be a complex one.

## 4.2 Technological and Econmic Study

Remote Manipulation has two main system groups. The first is the gesture capture and transmission portion. The second is the receiving system that performs the actual task. Both of these have their specific requirements but the goal of the RM system is to have the bulk of the equipment at the receiving end and as simple a setup at the transmission end.

By dividing the system into a simple and complex system, the transmission system could become no more complex that a video conferencing system. The benefit of a low cost transmission system is that control of the system could be done from almost anywhere that you can connect to the receiving system from. In the case of surgery, a specialist could then be anywhere instead of a specific clinic or hospital if there was an unexpected need for his or her skills.

On the receiving end however, there would be a significant technical burden on the development of the system. A great deal of technical proofs and knowledge would have to be included in the acting end of the system. Depending on the application this could be quite time consuming, costly and possibly not worth the investment. In the example of *Shaky Hand Surgery* the clinical costs and trials of implementing a functioning surgical system would likely be astronomical.

# 5 Moving Mannequin

## 5.1 Concept Generation, Design, Implementation

*Moving Mannequin* (MM) is a fun concept for retail environments. Rather than seeing how clothing looks on a static model in a shop window, a MM would mimic the movements of people standing in the shop window or passing by on the street.

The idea is simple, within reason, mimic the actions of a passer-by on the street outside the shop. This may range from a simple almost walk cycle or an attempt at recreating a dance being done by the person facing the mannequin. The MM may seem a simple way of engaging a passer-by outside a store, but in fact the relationship with the mannequin and the public goes a little further.

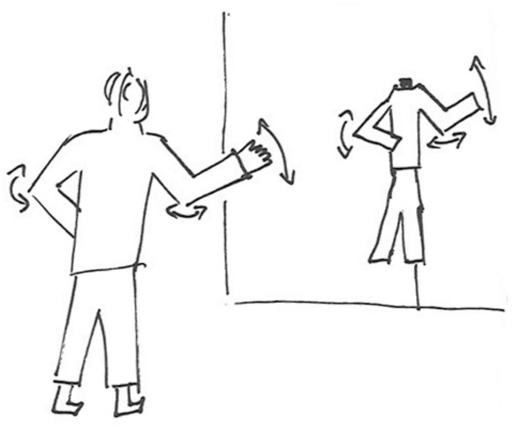

Figure 2.7 Moving Mannequin concept sketch

There are at least three possible groups of people involved with the MM system. The most obvious is the person acting something out and having a mannequin in a store window mirroring their actions. For some reason, having a system that mimics your actions, even in a limited way, seems to provide licence to the participants to push the bounds of what they may be willing to do in public. It may even inspire them to relax and have a little fun that they otherwise would not indulge in. The next group involved would be anyone on the street who is witnessing the action. There are probably countless unexpected spectacles that may be seen in the street. The third most obvious group is the store itself. With the rise of non-traditional marketing means, that now include anything other than print, television or radio, having a display that engages and draws people to a store longer may not only be fun, but profitable.

## 5.2 Technological and Econmic Study

The technology required for *Moving Mannequins* (MM) has two main component groups. The first is video based input system to capture user input. The second main component system would be the mannequin itself.

The implementation of the mannequin could be quite broad ranged. At one end of the spectrum, the mannequins could simply be augmented with the hardware required for movement: hinges, motors or linear actuators, etc. At the other end of the spectrum a humanoid robot, such as Palette from FlowerRobotics, could be used to implement the mannequin.

The economic impact of such a system in a storefront is difficult to judge. As with all advertising, the direct impact of one method of advertising is difficult to judge. With new trends in viral, gorilla and other 360 degree marketing campaigns the interest or revenue generated from non-traditional media advertising is even harder to judge.

Store owners may experience and increase in traffic directly outside their shops. With increased curiosity in the storefront, it is likely that some of the people outside the shop will enter the shop. Possibly some that would otherwise considered the shop. The true test would be to install the system and see if the number of overall consumers increased.

Depending on the complexity of the implementation, the value of the system can be measured by sales and revenue information compared with previous years and prior to the installation of a MM system.

# 6 Don’t Touch

## 6.1 Concept Generation, Design, Implementation

*Don’t touch* is concept to allow people to manipulate objects that they normally would not be able to handle. We’ve seen on many different product websites 3-dimenional (3D) models that can be tilted and/or rotated to provide different perspectives of mobile phones or many other gadgets. The idea behind the 3D models is that potential consumers can get a close up look at all the features a particular product may have to offer. What makes *Don’t Touch* different is that rather than manipulating a digital object, you get to manipulate a real object.

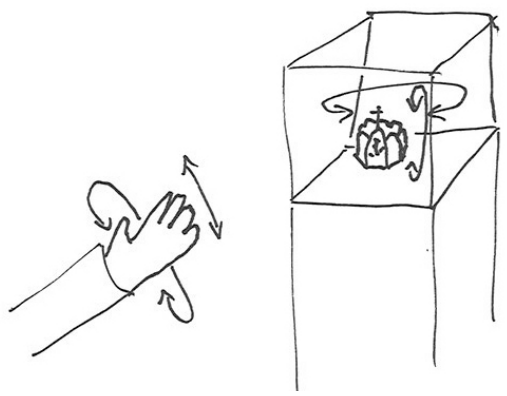

Figure 2.8 Don’t Touch concept sketch

You might find something like this in a number of different places. One place you might find this is in a shop window. Imagine standing outside a mobile phone shop and using your hands to roll around and rotate the latest mobile gadget or phone. Even if the store is not open you would have the opportunity to ‘handle’ the object and get to see it from all angles. Having this opportunity may increase the likelihood that a consumer will return later to purchase the item.

Another place you might find such a system could be in a museum. There are countless objects protected behind glass that we may never see but from the top down. This system could allow a visitor the deeper visual exploration of a piece – seeing all the different detail or how the light reflects off the object from different angles. Imaging being able to turn the crown jewels around and being able to examine them from all angles, something that you cannot do while they are locked in a glass display case.

Be it in a retail or museum environment *Don’t Touch* could allow people to handle and manipulate objects that for whatever reason they normally could not. This experience could create a deeper link with the user, while still maintaining control and security of the object *not quite* at hand.

## 6.2 Technological and Econmic Study

As mentioned earlier in this chapter this concept uses a video tracking system for input as well. The output portion of the system can be anything ranging from a simple gyroscope made of plastic to sit in a shop window to an entire enclosed system.

The technical implementation of either a standalone display, or an entirely enclosed display, is similar. Both would require a gyroscope system to move the display objects about. The value of the objects would likely determine the robustness of implementation. For a country's crown jewels, the system would likely be nearly indestructible.

Users of this system would more likely be museums and exhibition spaces. The benefit to a retailer of having a product that a consumer could manipulate outside of regular store hours may not justify the expense of the system when the same consumer would likely be able to physically handle the product during normal opening hours. Thus the real value would likely be for museums and galleries.

Deployment in a museum or gallery would provide visitors an experience opportunity they would not ordinarily have. This opportunity may increase interest in the exhibition and may draw more return visitors. Due to the possible expense of the system, the system could also be used as a draw for potential viewers. Obviously not all objects in a museum could be mounted in such a system, but items and artefacts of significance could be advertised as viewable with the system and could draw more visitors specifically for the opportunity to 'handle' the objects.

While the prevalence of hand gesture inputs is only an idea with no guarantee of successful adoption, as computing evolves, as interactive surfaces evolve, continuing to explore these concepts will

# 7 Final Evaluation and Analysis

Once the idea of multi-dimensional and multi-granular interface options was planted amongst my collegues at the Interaction Design Institute Ivrea, the number of usage scenarios started to expand rapidly. Staring with the simple ideas of object manipulation in 3 dimentional space, new navigation schemes emerged. Combined with the idea of context specific menus, notions of context specific controls that changed depending on the granular level of control being used evolved. With new and sometimes far-fetched hand gesture mapping ideas, the difficulty became how to most effectively exemplify the relationships your hands can have with a computing system and its applications.

Simple prototype applications can highlight the value of particular gestures or the relationships they can have with one another. Limiting one demonstration application to two hands controlling two values in an oversized calculator, demonstrates and immediate relationship of large and small scale movements that can affect a system. The object manipulation application, while not doing extraordinary things to the object itself, highlights the potentials of mapping several actions to one hand and the immediacy of action versus the shifting of focus to modal dialog windows. More of these prototype applications will undoubtedly uncover additional benefits and potentials from hand gesture inputs.

Presently the technology available to implement these systems is large and cumbersome. That does and should not limit the exploration of these concepts or possible interaction paradigms. In a non-graphical computing environment, arrow keys on the keyboard could have easily matched the use and effects of Douglas Englebart's first computer mouse, but the mouse, along with other its derivatives have become part of the interface normal of computing. Despite the absence of everyday graphical computer environments, the mouse and its models of use were developed. It is the hope of this thesis that with continued exploration and targeted prototyping of actions that a comprehensive and adaptable hands-on gesture input system can be developed.

The final evaluation and analysis at the time of this writing is that there is more to be explored and more to be discovered with hand gesture computing input systems.

# 8 Conclusion

With the boundaries between where you work, the keyboard below your eye line, and where your work is done, the monitor in your eye line, diminishing or ceasing to exist, the way we work will change. Using a table like surface or a multi-touch table computer will change the way we use indirect computing inferfaces. The keyboard no longer needs to be a physical object, it can be replaced with an onscreen representation. The mouse no longer needs to exist as your hands and fingers can do the pointing and tapping directly on the surface. The real extension to the WIMP paradigm will come with an evolution of multi-dimensional contexts and usages.

Undoubtedly if this input paradigm does find its place in tablet or table-top computing, it will take time to have a significant influence on the inertia of the keyboard and mouse input paradigm that dates from the early 1960s. It will likely only be once interface developers design from the start with full hand gestures in mind, that the richness and subtleties they can provide will truly realised.

# Sources

In this draft report at least these sources below, if not cited directly, have had some influence on the directions taken in this thesis investigation.

Coens, Laura. "The Invention of the Computer Mouse." Air Force Research Laboratory Technology Horizons. AFRL's Air Force Office of Scientific Research, Arlington, VA. 22 May 2006 <http://www.afrlhorizons.com/Briefs/Mar02/OSR0103.html>.

"Croquet Project." 19 Oct. 2005. Viewpoints Research Institute. 19 Oct. 2005 <http://www.opencroquet.org/About_Croquet/about.html>.

"The Devorak Keyboard." 24 Apr. 2002. SHIAR. 19 Oct. 2005 <http://www.shiar.org/happy/txts/dvorak.php>.

"Dvorak Simplified Keyboard." Wikipedia, the free encyclopedia. 19 Oct. 2005 <http://en.wikipedia.org/wiki/Dvorak_Simplified_Keyboard>.

"Introducing Archy." 19 Oct. 2005. Raskin Center. 19 Oct. 2005 <http://rchi.raskincenter.org/index.php?title=Home#Introducing_Archy>.

"Inroducing OmniGraffle 4." The Omni Group - Applications - OmniGraffle. 2005. The Omni Group. 21 Oct. 2005 <http://www.omnigroup.com/applications/omnigraffle/>.

"lg3d: Project Looking Glass." 2004. java.net. 19 Oct. 2005 <https://lg3d.dev.java.net/>.

Lieberman, Henry. "The Tyranny of Evaluation." 19 Oct. 2005 <http://web.media.mit.edu/~lieber/Misc/Tyranny-Evaluation.html>.

"Maltron keyboard." Wikipedia, the free encyclopedia. 19 Oct. 2005 <http://en.wikipedia.org/wiki/Maltron_keyboard>.

"'Minority Report' interface created for US military." 15 Apr. 2005. NewScientist.com. 19 Oct. 2005 <http://www.newscientist.com/article.ns?id=dn7271&feedId=online-news_rss20>.

Murch, Walter. *In the Blink of an Eye*. 2nd. Los Angeles: Silman-james Press, 2001.

Napier, John. *Hands*. Revised. Princeton, NJ: Princeton University Press, 1993.

"NASA Develops System To Computerize Silent, "Subvocal Speech"." 17 Mar. 2004. NASA. 19 Oct. 2005 <http://www.nasa.gov/home/hqnews/2004/mar/HQ_04093_subvocal_speech.html>.

Norman, Donald. Emotional Design. New York: Basic Books, 2004.

Spielberg, Steven .”Minority Report.” 2002. Film. <http://www.imdb.com/title/tt0181689/?fr=c2l0ZT1kZnx0dD0xfGZiPXV8cG49MHxrdz0xfHE9bWlub3JpdHkgcmVwb3J0fGZ0PTF8bXg9MjB8bG09NTAwfGNvPTF8aHRtbD0xfG5tPTE_;fc=1;ft=22;fm=1>.

Starner, Thad. Thad Starner Research. College of Computing, Georgia Institute of Technology. 19 Oct. 2005 <http://www.cc.gatech.edu/fac/Thad.Starner/031_research.htm>.

“Velotype.” Wikipedia, the free encyclopedia. 19 Oct. 2005 <http://en.wikipedia.org/wiki/Velotype>.

“Visio 2003 Home Page.” 2005. Microsfot Office Online. 21 Oct. 2005 <http://office.microsoft.com/en-us/FX010857981033.aspx>.

Wilson, Frank R.. *The Hand*. First Vinage Books. New York, NY: Vintage Books, 1999.

“xstroke: Full-screen Gesture Recognition for X.” xstroke. 19 Oct. 2005 <http://www.xstroke.org/usenix_2003/html/xstroke.html>.