10 mm

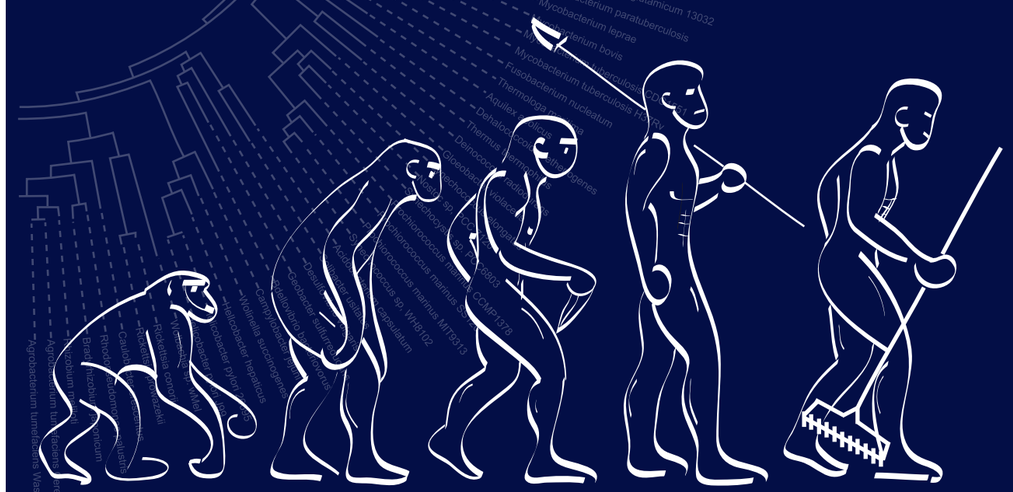
5 1

Gesture Interaction at a Distance

Wim Fikkert

Gesture Interaction at a Distance

Wim Fikkert



UITNODIGING

voor het bijwonen van de openbare verdediging van het proefschrift:

Gesture Interaction at a Distance

door

Wim Fikkert

op donderdag 11 maart 2010

Aanvang: 16:30 uur
Locatie: de Waaier, zaal 4
Universiteit Twente
Enschede

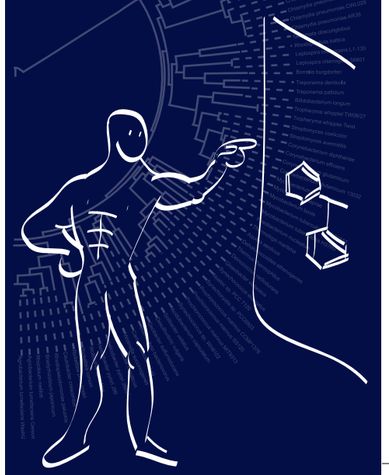
U bent vanaf 21:00 uur van harte welkom op het feest.

Locatie: 't Bölke
Molenstraat 6
Enschede

Voor meer informatie kunt u contact opnemen met:

Robert Moerland
r.moerland@xs4all.nl
0624409914

Ivo Swartjes
evaugh@gmail.com
0654280355



Gesture Interaction at a Distance

Wim Fikkert

PhD dissertation committee:

Chairman and Secretary:
Prof. dr. ir. A.J. Mouthaan, University of Twente, NL
Promotors:
Prof. dr. ir. A. Nijholt, Human Media Interaction, University of Twente, NL
Prof. dr. G.C. van der Veer, Open University Netherlands, NL
Assistant-promotor:
Dr. P.E. van der Vet, Human Media Interaction, University of Twente, NL
Opponents:
Prof. dr. ir. J. van Amerongen, Control Engineering, University of Twente, NL
Prof. dr. A. Eliëns, Control Engineering, University of Twente, NL
Dr.-Ing. S. Kopp, Sociable Agents Group, CITEC Cognitive Interaction Technology, Bielefeld University, DE
Prof. dr. J.A.M. Leunissen, Laboratory of Bioinformatics, Wageningen University and Research Centre, NL
Prof. dr. H. Reiterer, Human-Computer Interaction, Department of Computer & Information Science, University of Konstanz, DE
Dr. Z.M. Ruttkay, Human Media Interaction, University of Twente, NL
Paranymphs:
Dr. ir. ing. R.J. Moerland, Optical Sciences, University of Twente, NL
Ir. I.M.T. Swartjes, Human Media Interaction, University of Twente, NL

Human Media Interaction. The research reported in this thesis has been carried out at the Human Media Interaction (HMI) research group of the University of Twente, The Netherlands.

CTIT Dissertation Series No. 09-164. Center for Telematics and Information Technology (CTIT). P.O. Box 217, 7500 AE, Enschede, the Netherlands. ISSN: 1381-3617

NBIC Publication. This work is part of the BioRange program carried out by the Netherlands Bioinformatics Centre (NBIC), which is supported by a BSIK grant through the Netherlands Genomics Initiative (NGI). This thesis only reflects the author's views and funding agencies are not liable for any use that may be made of the information contained herein.

SIKS Dissertation Series No. 2010-07. The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

ISBN: 978-90-365-2973-0
ISSN: 1381-3617, number 09-164
DOI: 10.3990/1.9789036529730

GESTURE INTERACTION AT A DISTANCE

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. H. Brinksma,
on account of the decision of the graduation committee
to be publicly defended
on Thursday March 11, 2010 at 16:45 PM

by

Fredrik Willem Fikkert

born on January 17, 1981

in Vriezenveen, The Netherlands

This thesis has been approved by:

Promotors:

Prof. dr. ir. A. Nijholt
Prof. dr. G.C. van der Veer

Assistant-promotor:

Dr. P. E. van der Vet

ISBN: 978-90-365-2973-0

Dankwoord

Lieve Susan, ik heb je tijdens mijn promotie veel te vaak niet op de eerste plek gezet. Een plek die je wel verdient met alleen al alle geduld die je opbrengt tijdens de tien jaar dat ik inmiddels aan het studeren ben. Daarom wil ik graag mijn dankwoord beginnen met jou. Al die keren dat ik je de oren van de kop zeurde over nieuwe ideeën voor mijn proefschrift, steeds als ik weer eens de tijd vergat bij het op tijd thuis zijn voor het eten, dat ik weer eens weg ben naar een training of vergadering, dat je moest poseren om geschikt beeldmateriaal voor posters en papers te verkrijgen en de vele weekenden dat ik naar de UT ging om te schrijven aan mijn boekje zodat jij thuis in de zooi zat. Je bent er voor me op de momenten dat ik je nodig heb en daarvoor kan ik je nooit genoeg bedanken.

Dit dankwoord zal nu eerst even (heel kort) in het Engels zijn, daarna ga ik verder in het Nederlands. Dit betekent overigens niet dat er ook maar enige structuur in dit dankwoord zit, ik schrijf het gewoon op zoals het in me opkomt. Het volgende cliché volgt hier dan ook uit: ik ga mensen vergeten in dit dankwoord en speciaal voor iedereen die zich vergeten voelt (of die meer bedankt wil worden): bedankt!

Switching briefly to English. I would like to thank my committee for taking the time to read and exchange thoughts on gesture interaction at a distance. Job, Anton, Stefan, Jack, Harald and Zsofi: I feel really honored that you have taken this time and effort for me. I am looking forward to an interesting discussion at the defense ceremony. And switching back to Dutch already.

Om toch maar ergens te beginnen zal ik allereerst de mensen met wie ik de afgelopen jaren heb samengewerkt noemen. Mijn (co)promotoren zouden dan eigenlijk als eerste genoemd moeten worden maar daar komen we dadelijk wel aan toe. Eerst wil ik graag mijn kamergenoten bedanken die het met me hebben uit moeten houden. Sterker nog, ze zitten nog een jaar aan me vast. Ivo, Thijs, Herwin, Bart: ik vond het bijzonder gezellig en inspirerend om samen een kamer te delen bij HMI. Ivo voor je volstrekt onverwachtte plotwendingen; ook in gesprekken waar je niet eens actief aan deel leek te nemen, Herwin voor o.a. de gezellige trips naar Lissabon en Bielefeld, Bart voor de vele bierverhalen op de meest ongeschikte momenten, en Thijs voor de nuchtere vaderrol die je vaak speelde om het zooitje ongeregeld een beetje in het gareel te houden. Heren, onze vele lolletjes en inside jokes zullen me nog lang bijblijven. Naast deze heren hebben ook een paar dames tijdelijk of sporadisch een plekje geclaimd op onze kamer: Hanna en Yujia. Dames, ik vind het jammer dat jullie niet vaker en langer bij ons op de kamer hebben verbleven, dat zou denk ik ons puberale gedrag bij tijd en wijle in de kiem hebben gesmoord.

Dan ga ik nu graag door naar mijn projectgenoten. Ingo en Olga, bedankt voor

een leuke tijd. Ingo, van je nuchtere kijk op de zaak werd ik soms wanhopig wanneer je het probleem weer eens niet zei te zien. Ondanks dat we alledrie een totaal andere invulling hebben gegeven aan 'User interfaces for scientific collaboration' denk ik dat we veel aan elkaar hebben gehad. Ik voelde me in elk geval veelvuldig gedwongen om mijn ideeën beter of soms zelfs volledig opnieuw te doordenken om jullie duidelijk te maken wat ik van plan was. Dit geldt ook voor alle andere collega's bij HMI. Beste HMI'ers, ik ga jullie niet stuk voor stuk opnoemen want ik vind dat we samen een erg leuke en inspirerende werkomgeving neerzetten. De brede interesse in andermans onderzoek, de bereidheid om elkaar zonder meer te helpen met bijvoorbeeld deelnemen aan veelvuldige en soms saaie experimenten, doorlezen en bekritiseren van elkaars werk, maar ook de infrequente uitjes met elkaar dragen hier denk ik enorm aan bij. Bedankt allemaal dat jullie een steun en toeverlaat voor me waren tijdens mijn promotie. Ik hoop dat ik dit andersom ook voor jullie heb mogen en mag betekenen. Buiten HMI heb ik ook met diverse mensen mogen samenwerken en ideeën mogen uitwisselen: bij conferenties, workshops, werkbezoeken enzovoorts. Ook die collega's ben ik erkentelijk: Hanka, Gineke, Han, Timo, Werner en vele anderen.

Wel wil ik nog expliciet mijn promotoren, Anton en Gerrit, bedanken voor hun feedback op mijn onderzoek. Een leerpunt van me de afgelopen jaren was het inbedden van mijn werk in een grootschaligere opzet en ik werd daar door jullie dan ook veelvuldig daarop gewezen. Dit betekende dan weer meer nadenken maar het eindresultaat is daardoor flink beter geworden. Uiteraard heeft mijn copromoter Paul hier evenzo zijn steentje aan bijgedragen. Paul heeft twee kanten wat dat betreft: enthousiast en terughoudend. Enthousiast als je al stuiterend zijn kamer in huppelt met een nieuw en vet gaaf idee, terughoudend de paar minuten erna met een opmerking als 'zou je dat wel doen jongen?'. Bedankt voor je maatwerk begeleiding Paul, het heeft me veelvuldig ondersteund wanneer ik het nodig had en ik denk nog wel vaker op momenten wanneer ik er eigenlijk niet op zat te wachten.

Onderzoek doe je niet alleen, ook een promotietraject niet. Dit blijkt al uit het dankwoord richting mijn collega's. Op onze vakgroep lopen echter ook veel betrokken studenten rond. De (ex-)studenten die opdrachten bij mij hebben gedaan verdienen dan ook zeker een dankbetuiging. Ik noem 'mijn' afstudeerders en student assistenten met naam: Jorik, Jacobjob, Jeroen, Luke, Mario, Michiel en Marco. Heren, het was ook voor mij een leerzame ervaring jullie te begeleiden en ik stel het dan ook erg op prijs dat jullie dat hebben aangedurfd. De vele andere studenten (22 stuks) die ik heb mogen begeleiden met individuele of groepsopdrachten hebben me geholpen om ook mijn eigen ideeën beter te vormen waarvoor ik ook hen zeer erkentelijk ben.

Hobbies dan. Ik heb er maar twee: onderwaterhockey en duiken bij ZPV Piranha. Echter weet ik me hier zo betrokken bij te voelen dat de meeste van mijn vrije tijd daaraan opgaat. Gelukkig maken de vele leuke mensen die ik daar heb ontmoet heel veel goed. Sterker nog, door hen voel ik me zo betrokken. Samen weten we er een bijzonder leuke club van te maken. Bedankt dus, lieve Piranha's, voor een onvergetelijke tijd in en rondom het water! Onderwaterhockey is mijn uitlaatklep: zodra mijn hoofd onder water komt, kom ik tot rust. Rust ondanks dat ik me feitelijk

helemaal kapot zwem. Met name Marco, Eike, Wouter, Chris, Sandor, Harry, Steef en Ivo krijgen me zover om een betere speler te willen worden, inmiddels tot op het hoogste niveau in Nederland toe. Jullie enthousiasme voor de sport zweept me op en daarvoor ben ik jullie heel dankbaar. Duiken daarentegen is een heel serieuze bezigheid (geen sport). Een veelgebruikte slogan is 'duikers zijn gelukkiger'. Ja, daar kan ik me in de drie jaar dat ik nu duik goed in vinden. Nederland, Gozo, Zwitserland en Egypte zijn mooi onder water en dat is iets dat veel te weinig mensen weten en waarderen. Echter is dat niet hetgeen waardoor duikers gelukkiger zijn vind ik. Het zit hem in de gezelligheid rondom het duiken. Charlotte, Saskia, Kirsten, Julia, Wendy, Bas, Robbert, Robert, Marco, Erik, Jeroen, Arjan, Ingmar, Michel, Rob, Siebren, Timo, Martin en vele anderen: jullie maken duiken een gave ervaring die vrij letterlijk almaar naar meer smaakt. Bubbels bubbels bubbels!

Zeker tijdens het laatste jaar van je promotie heb je maar weinig tijd voor je vrienden. Dat is zo ongeveer vanaf het moment wanneer je alles wat je nog moet doen een beetje kunt overzien en vervolgens bijna in huilen uitbarst van die hoeveelheid werk. Tel daar die tijdrovende hobbies bij op en ik ben werkelijk stomverbaasd dat ik zulke goede vrienden heb. Jelly, Femke, Marten, Reinier, Martijn en Hans: ik vind het tof dat we elkaar altijd weer weten te vinden. Ik hoop dat we de komende tijd weer meer van elkaar zullen zien!

Mijn familie komt dit keer op de laatste plaats. Pap, mam, ik hoop dat jullie eindelijk genoeg hebben van die alsmaar studerende zoon van jullie. Ik wel namelijk. Dit is de derde en laatste keer dat jullie een schijnbaar onnavolgbaar verhaal moeten aanhoren. Jullie hebben Marieke en mij altijd gestimuleerd om het beste te worden dat we kunnen zijn. Dat dit tot gevolg heeft gehad dat mijn zusje piloot is geworden en dat ik inmiddels een betaald warhoofd ben hadden jullie vast niet verwacht. Bedankt pap, bedankt mam. Ik was nooit het warhoofd geworden die ik nu ben zonder jullie steun, vertrouwen en liefde. Bedankt Marieke, dat je er voor me bent als ik je echt nodig heb. De vele nutteloze conflicten van toen hebben inmiddels plaats gemaakt voor een, vind ik, fijne zus-broer relatie. Bedankt Nick dat je je zo goed over mijn zusje ontfermt.

Mijn proef van bekwaming in de vorm van dit proefschrift is nu voltooid. Tijdens het doen van de experimenten en het rapporteren daarvan in de dit proefschrift heb ik ontdekt dat ik onderzoek doen heel leuk vind, vooral in een omgeving met mensen die net zo enthousiast worden van hun ding als ik van het mijne. Zo steek ik in elkaar: ik voed me met de passie van anderen. Ik hoop dan ook dat ik in de toekomst meer van zulke mensen mag ontmoeten met wie ik knappe staaltjes werk neer mag zetten, zowel professioneel als privé.

Wim Fikkert
Enschede, 1 februari 2010

voor Simay & Susan

Summary

The aim of this work is to explore, from a perspective of human behavior, which gestures are suited to control large display surfaces from a short distance away; why that is so; and, equally important, how such an interface can be made a reality. A well-known example of the type of interface that is the focus in this thesis is portrayed in the science fiction movie 'Minority Report'. The lead character of this movie uses hand gestures such as pointing, picking-up and throwing-away to interact with a wall-sized display in a believable way. Believable, because the gestures are familiar from everyday life and because the interface responds predictably.

Although only fictional in this movie, such gesture-based interfaces can, when realized, be applied in any environment that is equipped with large display surfaces. For example, in a laboratory for analyzing and interpreting large data sets; in interactive shopping windows to casually browse a product list; and in the operating room to easily access a patient's MRI scans. The common denominator is that the user cannot or may not touch the display: the interaction occurs at arms-length and larger distances.

Hand and arm movements are the gestures that computer systems interpret in this thesis. The users can control the large display, and its contents, directly with their hands through acts similar to those in 'Minority Report'. The control is gained through explicitly issuing commands to the system through gesturing. After defining the elementary commands in such an interface (Chapter 2), we index existing approaches to build gesture-based interfaces (Chapter 3) and, more precisely, the gesture sets that have been used in these interfaces. Meticulous investigation of which gestures are suited for issuing these elementary commands, *and why*, then follows.

In a Wizard of Oz setting, we explore the gestures that otherwise uninstructed users make when asked to issue a command through gesturing alone (Chapter 4). By gesturing as they see fit, users pan and zoom a map of the local topology of our university. Our observations show that users apply the same idiosyncratic gesture for each command with a great deal of similarity between users. Also, gestures are explicitly started and ended by changing the hand shape from rest to tensed and back again. Users really believed that they were in actual control of the display; immersed in the interaction that they found believable.

This consensus in the observed gestures is explored with an online questionnaire (Chapter 5) filled out by a hundred users from multiple western countries. User ratings of video prototyped interactions through gesturing show that there is significant preference for certain gesture-command pairs. In addition, some gestures are

preferably reused in a different context or system state to improve understanding and predicting of the system's responses. These results are validated in another (partial) Wizard of Oz setting (Chapter 6) where the users experience what it feels like to issue commands with the proposed gestures. The ratings in each investigated condition were similar, with minor differences that are mostly caused by physical comfort, or lack thereof, while gesturing. Our findings were influenced profoundly by both traditional WIMP-style interfaces and recent mainstream multi-touch interfaces that swayed our participants' preference towards some gestures.

To consolidate our previous findings, we designed, built and evaluated a gesture interface with which the user can interact with 3D and 2D visualizations of biochemical structures on a wall-sized display (Chapter 7). This prototype uses lasers for pointing, one for each hand, and small buttons attached to the fingers for issuing commands. The preferred gestures define the precise layout of these buttons on the hand. Again, we found that our participants preferred to interact with the least amount of effort and with the highest comfort possible. There was little variation between users in the shape of the gestures that they preferred: tapping the thumb on one of the other fingers was the prevalent gesture to indicate the beginning and ending of a command: it mimicked pressing a button.

When taking a human perspective on gestures suited to issue commands to large-display interfaces, it is possible to formulate a set of intuitive gestures that comes naturally to its users. The gestures are learned and remembered with ease. In addition, it is comfortable to perform these gestures, also when interacting for longer periods of time. We observe in our line of research that technological developments that reach mainstream distribution in the public domain influence the perception of 'intuitive' and 'natural' in the end-users. The best example of this is perhaps the influence of the indoctrination over the past four decades that the keyboard-and-mouse interface has had on the public's notion of human-computer interaction. More recent examples include the Nintendo Wii and the Apple iPhone. We, as the interface designers of future intelligent environments, are very much dependent on this notion. That is, if we wish to have gesture-based interfaces succeed in providing easy to use, intuitive interaction with the pervasive large display surfaces in these environments. The gestures that are described in this thesis are an important part of those interfaces.

Samenvatting

Het doel van dit werk is om vanuit het oogpunt van menselijk gedrag te ontdekken welke gebaren geschikt zijn om grote digitale oppervlakken te bedienen vanaf een korte afstand; waarom dat zo is; en, even zo belangrijk, hoe een op gebaren gebaseerde interface werkelijkheid gemaakt kan worden. Een bekend voorbeeld van het type interface dat de focus is in dit proefschrift is te vinden in de science fiction film 'Minority Report'. Het hoofdpersonage gebruikt handgebaren zoals wijzen, oppakken en weggooien om op een geloofwaardige manier te interacteren met een computerscherm ter grootte van een hele muur. Geloofwaardig, omdat de gebaren herkenbaar zijn uit het dagelijks leven en omdat de interface reageert op een voorspelbare manier.

Alhoewel de interface in deze film slechts fictief is zullen gebareninterfaces, wanneer ze gerealiseerd zijn, worden toegepast in een omgeving die is uitgerust met grote computerschermen. Bijvoorbeeld, in laboratoria om grote data sets te analyseren en te interpreteren; in digitale etalages om in het voorbijgaan nieuwe productinformatie te bekijken; in operatiekamers om snel en gemakkelijk toegang te krijgen tot MRI-scans van de patiënt; en in publiekelijk toegankelijke kunsttentoonstellingen waar een interactieve creatieve ervaring ondergaan kan worden. De gemene deler is hierbij dat de gebruiker het computerscherm niet mag of kan aanraken: de interactie vindt plaats op armslengte en grotere afstanden.

Hand- en armbewegingen zijn de gebaren die computersystemen interpreteren in dit proefschrift. De gebruikers kunnen het grote scherm, en de visualisaties daarop, rechtstreeks bedienen door handelingen met hun handen die lijken op de handelingen in 'Minority Report'. De controle wordt verkregen door expliciet commando's te geven aan het systeem door middel van gebaren. Nadat er een set aan elementaire commando's in een zodanige interface is gedefinieerd (Hoofdstuk 2) geven we een overzicht van bestaande manieren om een gebareninterface te bouwen (Hoofdstuk 3) en, preciezer, de gebarensets die daarin gebruikt worden. Nauwgezette bestudering van geschikte gebaren voor de elementaire commando's, *en* de redenen die daaraan ten grondslag liggen, beslaan de rest van dit proefschrift.

Met een Tovenaar van Oz proefopstelling hebben we spontane gebaren bestudeerd. Gebruikers werden gevraagd om een commando te geven door met hun handen te gebaren waarbij ze niet verteld werd hoe dat gedaan moest worden (Hoofdstuk 4). Door gebaren te maken zoals zij die nuttig achtten konden deze gebruikers een stafkaart van Twente verplaatsen, vergroten en verkleinen op een groot computerscherm. Onze observaties laten zien dat gebruikers een gebaar kiezen voor elk commando, dat ze aan hun keuze vasthouden en dat hun keuze grotendeels gelijk

is aan die van andere gebruikers. Daarbij vonden we ook dat gebruikers expliciet hun handvorm veranderden bij het begin van een gebaar en dat ze hun hand ontspanden aan het eind van een gebaar. Bovendien verkeerden de gebruikers in de waan dat zij daadwerklijk de controle over het scherm voerden.

Deze consensus in geobserveerde gebaren hebben we verder bestudeerd met een online vragenlijst (Hoofdstuk 5) die door honderd gebruikers uit verscheidene Westerse landen is ingevuld. De gebruikersscores van videoprototypes van gebareninteracties laten zien dat er een significante voorkeur is voor specifieke combinaties van gebaar en commando. Daar komt bij dat sommige gebaren bij voorkeur hergebruikt worden in een andere context binnen het systeem zodat het begrijpen en voorspellen van de systeemreacties vergemakkelijkt wordt. Deze resultaten zijn gevalideerd in een nieuwe Tovenaar van Oz proefopstelling (Hoofdstuk 6) waarin we gebruikers hebben laten ervaren hoe het *echt* is om middels de voorgestelde gebaren commando's te geven aan een groot computerscherm. De scores van deze validatiecondities waren gelijk, met slechts kleine verschillen die werden veroorzaakt door fysiek comfort of het gebrek daaraan tijdens het gebaren. De voorkeuren voor bepaalde gebaren werden sterk beïnvloed door zowel traditionele WIMP en recentere multi-touch interfaces.

Om de verkozen gebaren van dit grootschalige onderzoek te consolideren hebben we een gebaren interface ontworpen, gebouwd en geëvalueerd. De gebruikers konden interacteren met zowel 3D als 2D visualisaties van biochemische structuren op een scherm ter grootte van een hele muur (Hoofdstuk 7). Dit prototype gebruikt lasers voor aanwijzen, een voor elke hand, en kleine knopjes die op de vingers zijn geplaatst om commando's te kunnen geven. Wederom vonden we bewijzen dat onze gebruikers de voorkeur hebben voor interacties die de minste moeite kosten en die hen het meeste comfort bieden. Er was weinig variatie tussen gebruikers in de vorm van voorkeursgebaren: het tikken van de duim op een van de andere vingers was het meest voorkomende gebaar waarmee het begin en eind van een commando werd gekenmerkt. Dit gebaar lijkt op het drukken op een knop.

Wanneer we vanuit het menselijk perspectief kijken naar geschikte gebaren om commando's mee te geven naar grote computerscherm interfaces is het mogelijk om set van intuïtieve gebaren te formuleren die natuurlijk overkomen op de gebruiker. Deze gebaren worden met gemak geleerd en onthouden. Het is bovendien comfortabel om op deze manier te gebaren, ook wanneer de interactie langere tijd duurt. We hebben in onze lijn van onderzoek recente technologische ontwikkelingen geobserveerd die de notie van 'intutief' en 'natuurlijk' flink hebben beïnvloed bij onze gebruikers. Het beste voorbeeld is wellicht de indoctrinatie gedurende de afgelopen vier decennia door muis en toetsenbord maar ook meer recentere ontwikkelingen zoals Nintendo's Wii en Apple's iPhone. Wij, als interface-ontwerpers van toekomstige intelligente omgevingen, zijn heel afhankelijk van die publieke notie. Als we gebareninterfaces willen laten slagen in zulke omgevingen zullen we een gemakkelijk te gebruiken, intuïtieve interactie moeten ontwerpen en bouwen met de grote computerschermen die we overal om ons heen zullen gaan aantreffen. De gebaren die in dit proefschrift worden beschreven, zijn daar een belangrijk onderdeel van.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> "A computer terminal is not some clunky old television with a typewriter in front of it. It is an interface where the mind and body can connect with the universe and move bits of it about."
>
> **Douglas Adams**
> British writer, 1952–2001 – Mostly Harmless, Picador, 2002, pp.86–87

1.1 What this thesis is about

John Anderton stands calmly facing an empty wall from some two meters away, staring, arms crossed behind his back. He drops his hands by his side and, then, as he lifts them, palms upwards, the entire wall comes to life with pictures of a crime that John and his police colleagues are trying to solve. Through simple, familiar acts with his hands, such as pointing, grabbing and throwing away, he is able to sift through data such as pictures of the crime-scene and personal details of both the victim and the suspect to get clues for *preventing* the crime: Minority Report is a science fiction movie set in the mid 21st century, after all. The acts—gestures—that John uses in this setting—human-computer interaction (HCI)—is what this thesis is all about. We look at these gestures from a human perspective in this thesis, wondering which gestures are suited to interact with these large displays, why that is so and how an interface such as the one portrayed in Minority Report can be made a reality.

Large displays will not be found solely in John's fictional crime-lab. Future homes [200], offices [148], schools [125] and other public environments [214] will be equipped with displays that can be found anywhere: from newspapers lying around to clothing, furniture, the floor and the walls [67]. My thesis focuses on the latter type: physically large display surfaces that can display a lot of information simultaneously for the environments' inhabitants to interact with. Humans that interact with these displays will do so in one of two ways according to trends in HCI research [16]. First, human-like communication, for example, by conversing with a

digitized human, allows users to operate the computer in ways that mimic a dialogue with another person [179]. Second, a more direct way of interacting is the result of explicit command-giving [190]. The latter approach has emerged over the past decades since the advent of personal computers and is typically referred to as the WIMP metaphor: Windows-Icons-Menus-Pointing [207]. I have focused on the latter of these two types: explicit command-giving. The difference here is that the direct interaction occurs through the hands gesturing.

The distance to the display has an important influence on the interaction with it. John stands out of arms-reach of the display, making it impossible to touch it, while, at the same time, allowing spectators—John's fellow policemen—to view everything that he is doing. When he is standing at arms-length of the display, John is in the action zone. There, he can, but is not required to, touch the display. John can move from the action zone to the negotiation zone where he can no longer touch the display. However, when John is standing in the negotiation zone, his fellow policemen can observe, and possibly respond to, John's interactions [60; 72]. While John is interacting, his colleagues are standing in the reflection zone: they do not have the immediate intent to act. When influencing John's interaction, for example, with a comment, these colleagues remain in the reflection zone. If they, however, directly interact with the display, as John does, they move towards the negotiation zone. Because spectators can view John's interaction, privacy issues might arise [102]. In this thesis, I focus on the negotiation and reflection scales of interaction where the user is unable or not permitted to touch the display.

1.2 Origins of this thesis

This work is part of the BioRange program carried out by the Netherlands Bioinformatics Centre (NBIC), which is supported by a Bsik grant through the Netherlands Genomics Initiative (NGI). The BioRange project formally started in 2005 to promote the collaboration of Dutch research institutes and universities active in the life science domain. This thesis is part of the subproject 4.2.1 "User interfaces for scientific collaboration" in the context of virtual laboratories for *e*-Science.

The work described in this thesis was done at the Human Media Interaction (HMI) group of the University of Twente, the Netherlands. There, we look into ways that the computer can operate in every day life as universal media machines that present multi-media information and as communication devices that connect people. The interface is the central topic at HMI. At HMI we study various aspects of the interface which we address through speech, computer vision, virtual agents, storytelling, games and, in this thesis, gesture interfaces.

1.3 Application possibilities

Gesture interfaces can be applied in various display-rich environments. We describe four examples to demonstrate where such an interface can be applied and in what

way. The common denominator is that the display is accessible to several people simultaneously, even though it might be controlled by just one of them. Note that each separate interaction with these displays do not last for extended periods, say, longer than 10 minutes.

1.3.1 *e*-BioLab

At the University of Amsterdam, life scientists have built a display-rich meeting room to aid the analysis of their microarray[1] experiments. The enormous amounts of data generated in these and similar experiments have shifted the bottleneck of life science research from data generation to the storage, analysis and interpretation of these data. This process requires the analysis of hundreds of scatter plots that result from the statistical analysis of microarray scans. By projecting these data simultaneously on large interactive surfaces in the *e*-BioLab, the life scientists can make sense of their data, from both the generated overview and the details of each individual diagram [176].

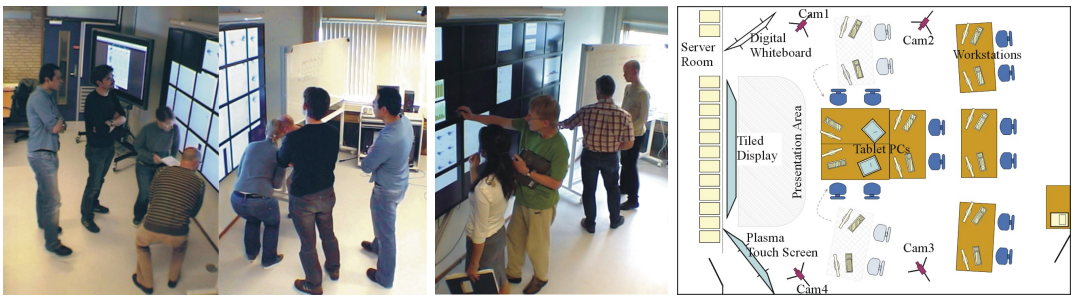


Figure 1.1: The Amsterdam *e*-BioLab is a display-rich meeting room that targets scientific discussions. (a) and (b) Users walk up to the display and point out the plots and other project results that they are discussing while (c) large physical display surfaces facilitate the simultaneous display of large numbers of project results. Note the cameras for behavior observation.

Multidisciplinary teams use the *e*-BioLab to study genome expression profiles while aiming, for example, to develop new medicines. These teams discuss their project results in front of the display, see Figure 1.1. It is important to note that currently it is not possible to control the displays in the *e*-BioLab directly; an operator (not in these photos) manages the display contents based on explicit user requests. By offering a direct means of selecting, manipulating and correlating pieces of data on the display, researchers are handed a true tool for furthering their research process.

[1]Microarrays are a recent technology in biological research with which the expression levels of thousands of genes can be investigated simultaneously [185].

1.3.2 Shopping area

In an average shopping area, display windows increasingly try to catch the eye of passers-by through movement, for example, with videos of products on sale. Through various sensors, it is possible to look at the passers-by and to tune the videos to the behavior of the humans in front of the window. For example, if someone stops in front of the display, an advertised product might be shown in more detail. The time that users stand in front of a shop is typically brief, varying from a short glance to a short stop for a minute or two [139]. In addition, privacy issues, caused by the display being open to other onlookers, discourage users to interact extensively [15]. The interaction should convey as much product information as possible to the customer-to-be, requiring an easy-to-use, non-invasive interface. By moving between interaction zones—action, negotiation and reflection—the privacy concern might be alleviated by adjusting the amount of detail shown to the distance the user is standing from the display [214]. The availability of mobile phones and the variety of motion sensors they contain nowadays make it conceivable to link these devices to the display window for interacting with and exchanging product information [213, Ch.4].

1.3.3 Operating room

In some environments the user is not allowed to touch a display. A prime example is the operating room where the hands must remain sterile throughout the entire procedure: keyboards and mice have proven to be a common means for spreading infections in intensive care units [217]. Touch-based interfaces have even been dubbed 'the most evil technology in modern computing' because of their potential to spread disease [2]. During surgical procedures doctors require access to specific patient information, for example, MRI, CT and X-ray images. In current situations, the surgeon requests this information from other medical staff at a main control wall. A gesture interface can facilitate navigation, selection and manipulation of images by the surgeon who does then not need to leave the patient to access the main control wall. Note that the operating room then has to become a display-rich environment. Time-critical tasks during surgery are in such a scenario supported by facilitating information access through a gesture-interface in the sterile operating room [216].

1.3.4 Interactive public art

Creative environments are another application area for gesture interfaces. By surrounding users with display surfaces, it becomes possible to immerse those users in a virtual world. By pointing out objects of interest in this virtual world, users can navigate in and interact with the world with ease [114]. The virtual world does not have to be a realistic projection but it can, instead, be artistic in nature. When left to their own devices, individual users and user groups tend to explore such interactive

[2] http://blogs.zdnet.com/igeneration/?p=776, October 12th, 2009.

systems for their own enjoyment and creative self-perception by producing artistic interactions [62]. Existing art works in museums can also be made interactive for the user to experience [162] and indeed, whole museums might benefit from the interactivity that gesture-interfaces have to offer [117].

1.4 What this thesis is *not* about

The possibilities for doing research into gesture interfaces are limitless. In this work we aim to explore gesture-based interaction with large displays that cannot or may not be touched. We now define the boundaries of this work by describing what this thesis is *not* about. Summarizing: this thesis is not about computer vision, touch-sensitive surfaces, sign-language, indirect or deictic interfaces.

Computer vision is considered to be the least invasive and most promising method for looking at users gesturing for interpretation [168]. Algorithms are being developed that detect, track, recognize and interpret the shape and movements of human fingers, hands and arms from camera images and image sequences. We use techniques such as computer vision rather than develop our own automated method for looking at gestures. The promise of this research field has been, over the past decade, to make robust algorithms for analyzing and interpreting camera images. However, the current state-of-the-art is still not mature enough for robust detection of human gesturing in real-world surroundings.

The interactions we explore in this thesis focus on direct communication between user and system. Directly interacting with the display means that there is a direct connection between the user gesturing and the responses of the system. In contrast, indirect interactions, for which the mouse is perhaps the best known example, separate the input from the feedback spatially [52]. There is no looking up from a tablet to the large display to see what the system's response is to your actions.

Touch-sensitive surfaces, multi-touch technology in particular, have taken flight during the formation of this work [73]. These interfaces are perhaps the ultimate direct interface and we looked into them extensively. As a result, they have contributed significantly to the formation of the ideas discussed in this thesis. However, we excluded touch-sensitive surfaces from this thesis because they have more to do with the design of the graphical interface and the interplay between that interface and the act of touching it by the user.

A large part of research into gestures in HCI is being focused on sign-language systems, for example, addressing sign-language education programs for young children [202]. It involves machine analysis and understanding of human action and behavior, tracking and segmentation of human motion analysis, and gesture recognition [155]. Signed languages are as rich and complex as any spoken language with complex spatial grammars that convey meaning. The interfaces that look at sign-language are based upon the predefined signs and grammatical structure in which they occur.

The first gesture interface, Bolt's "put that there" system [11], showed the po-

tential for combining speech with gesturing in HCI. Such multimodal systems build upon natural human dialogue that combines speech and gesturing [16]. Gestures in these systems are not a separate entity used for issuing commands; they rather disambiguate and elaborate on spoken commands.

1.5 Contributions of this thesis

The aim of this work is to explore, from the perspective of human behavior, which gestures are suited to control large displays, why that is so and, equally important, how an interface such as the one portrayed in Minority Report can be made a reality. This thesis makes several contributions; we contribute:

- A four-state model of interface states and state transitions from a user's perspective (**Chapter 2**). The state transitions represent commands that, depending on the interface itself, can be suitably issued with, in the case of this thesis, gestures;
- Guidelines for the design, implementation and evaluation of gesture interfaces that follow from insights gained in an experiment with uninstructed gesturing for command-giving (**Chapter 4**) and experiments that explored which gestures are intuitive for HCI and why that is so (**Chapters 5** and **6**);
- Further evidence that online video clips can be used to instruct gesturers how to interact through gestures with large display interfaces. Instructions through actively performing the gestures differs only marginally from passive, online instructions (**Chapters 5** and **6**);
- A prototype of a gesture interface that uses a validated gesture-set for issuing elementary commands to a system with a large display interface from a distance beyond arms-length (**Chapter 7**).

1.6 Published as

Parts of this dissertation have been published before. In this work we elaborate and complement these publications. These publications are:

[43] W. Fikkert, M. D'Ambros, T. Bierz, and T. Jankun-Kelly. Interacting with Visualizations. In: A. Kerren, A. Ebert, and J. Meyer, eds., *Human-Centered Visualization Environments*, vol. 4417/2007 of *Lecture Notes in Computer Science, GI-Dagstuhl Seminar 3*, pp. 77-162. Springer Verlag: 2007.

[49] W. Fikkert, P. van der Vet, H. Rauwerda, T. Breit, and A. Nijholt. A Natural Gesture Repertoire for Cooperative Large Display Interaction. In: *Advances in Gesture-Based Human-Computer Interaction and Simulation*, vol. 5085/2009 of *Lecture Notes on Computer Science*, chap. 22, pp. 199-204. Springer Berlin / Heidelberg: 2009.

[46] W. Fikkert, N. Hoeijmakers, P. van der Vet, and A. Nijholt. Navigating a Maze with Balance Board and Wiimote. In: *The 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN '09)*, vol. 9 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 187-192: 2009.

[48] W. Fikkert, P. van der Vet, and A. Nijholt. Gestures for Large Display Control. In: *Gesture in Embodied Communication and Human-Computer Interaction*, vol. 5934/2009 of *Lecture Notes in Computer Science*, p. 12. Springer, Berlin: 2009, in press.

Other publications that have contributed in the formation of this dissertation and the ideas that it contains, but that have not been included in this work are:

[209] P. van der Vet, O. Kulyk, I. Wassink, W. Fikkert, H. Rauwerda, B. van Dijk, G. van der Veer, T. Breit, and A. Nijholt. Smart Environments for Collaborative Design, Implementation, and Interpretation of Scientific Experiments. In: *Workshop on AI for Human Computing (AI4HC)*, vol. 20 of *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 79-86. AAAI Press: 2007.

[47] W. Fikkert, H. van der Kooij, Z. Ruttkay, and H. van Welbergen. Measuring Behavior using Motion Capture. In: A. Spink, M. Ballintijn, N. Bogers, F. Grieco, L. Loijens, L. Noldus, G. Smit, and P. Zimmerman, eds., *Proceedings of Measuring Behavior 2008, 6th International Conference on Methods and Techniques in Behavioral Research*, p. 13. Noldus, Maastricht, The Netherlands: 2008.

[44] W. Fikkert, M. Hakvoort, P. van der Vet, and A. Nijholt. Experiences with interactive multi-touch tables. In: *The 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN '09)*, vol. 9 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 193-200. Springer Berlin Heidelberg: 2009.

[45] W. Fikkert, M. Hakvoort, P. van der Vet, and A. Nijholt. FeelSound: Collaborative Composing of Acoustic Music. In: *Proceedings of the 6th International Conference on Advances in Computer Entertainment Technology (ACE '09)*, pp. 294-297. ACM, Athens, Greece: 2009.

1.7 Dissertation structure

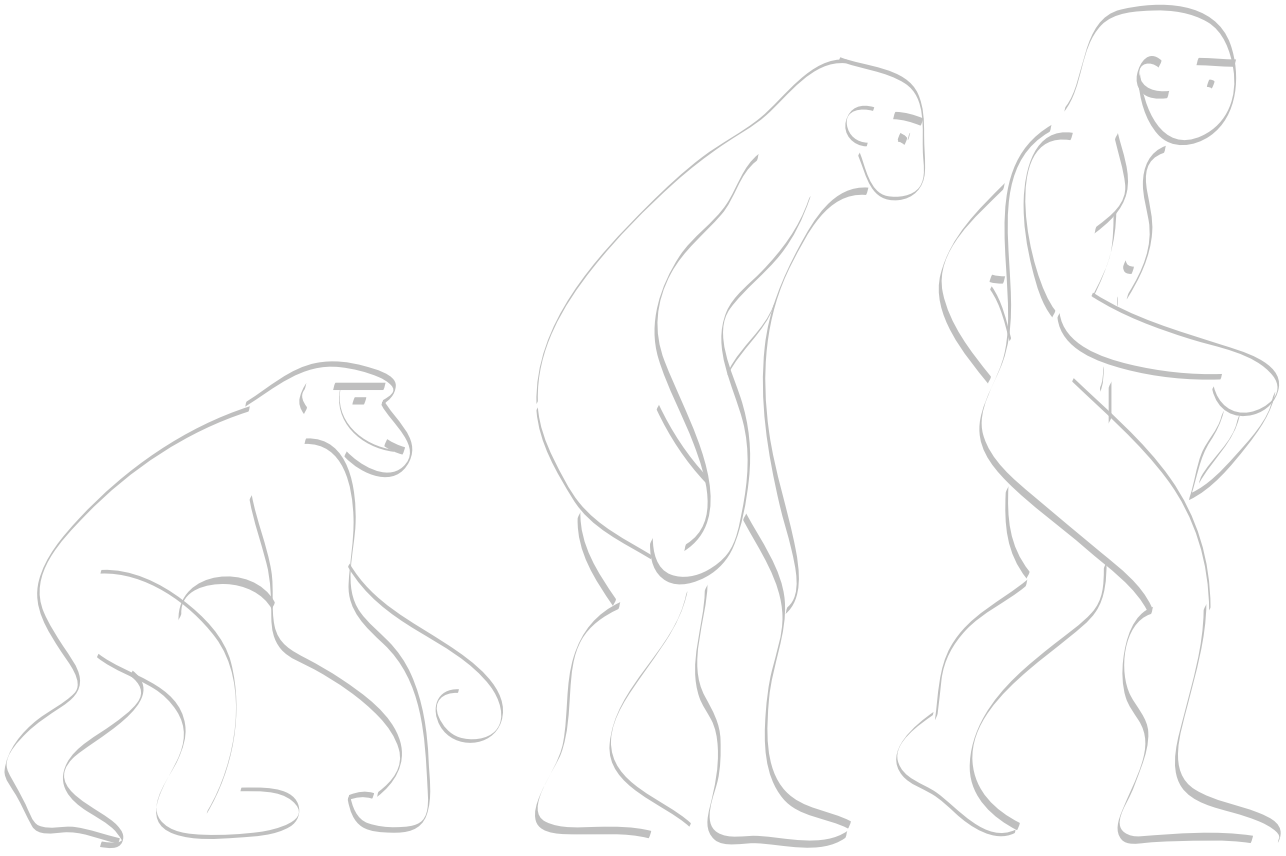
This dissertation is divided into three parts. Part I on related work describes in **Chapter 2** what human-computer interaction is, which elementary interface tasks can be distinguished in a technological environment and how such an environment can sense its human inhabitants gesturing. In **Chapter 3** we describe in detail what gestures are, how they can be categorized and how they have been applied in HCI.

Part II describes four experiments in which the human perspective on intuitive gesturing for command-giving is meticulously investigated. The first experiment is described in **Chapter 4** where we asked uninstructed users to issue commands through gestures. **Chapter 5** evaluates a large set of potentially useful gestures for issuing elementary interface commands with a large-scale online questionnaire. In **Chapter 6** we validate those findings in two smaller experiment conditions with a prototype interface. **Chapter 7** puts together all previous findings in a fully working, gesture large display interface with a wall-sized display.

We wrap up this thesis in Part III with conclusions based on our findings in **Chapter 8** and with a discussion and future vision that are based on the implications of our findings.

Part I

Related Work



Chapter 2

HCI and Gestures

> "In a computer controlled environment one wants to use the human hand to perform tasks that mimic both the natural use of the hand as a manipulator, and its use in human-machine communication."
>
> **Vladimir Pavlovic, Rajeev Sharma and Thomas Huang**
> [166, p.679] *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19 (7): 677-695: 1997.

The previous chapter described the motivation, application possibilities, boundaries and structure of this work. This part on related work starts with a sketch in Section 2.1 of the multidisciplinary human-computer interaction (HCI) research field, how gesture-based interfaces are a part therein and it places the work presented in this thesis in a HCI context. In Section 2.2 we describe the type of interface that we address in this thesis in more detail than we did in Chapter 1. We then explore the tasks that lie at the heart of a reactive interface that is controlled through the hands gesturing for explicit command-giving in Section 2.3. As this is directly dependent on the sensors that are used in these interfaces, we also give an overview of various input and output modalities suited for gesture-based interfaces in Section 2.4. The following chapter in this part (Chapter 3) then defines what gestures are, how they can be categorized and how gestures have been used in HCI.

2.1 The HCI field

The human-computer interaction (HCI) field studies the relationship between humans and their technological environment [12]. In this process, researchers develop diverse interaction solutions with which the humans and their environment can exchange information. Studying these interactions requires knowledge and inspiration from multiple research fields: social and engineering sciences in addition to design, see Figure 2.1a. The social sciences bring, amongst other things, sociology, psychology, communications theory and anthropology to the HCI table. The engineering sciences contribute computer science, electrical and mechanical engineering, physics, and information representation. The third research field in HCI provides

knowledge on architecture, graphic design and industrial design. The coming together of these three disciplines has, gradually, lead to a greater understanding of the workings and ways to further existing and new paradigms of human-computer interactions [221].

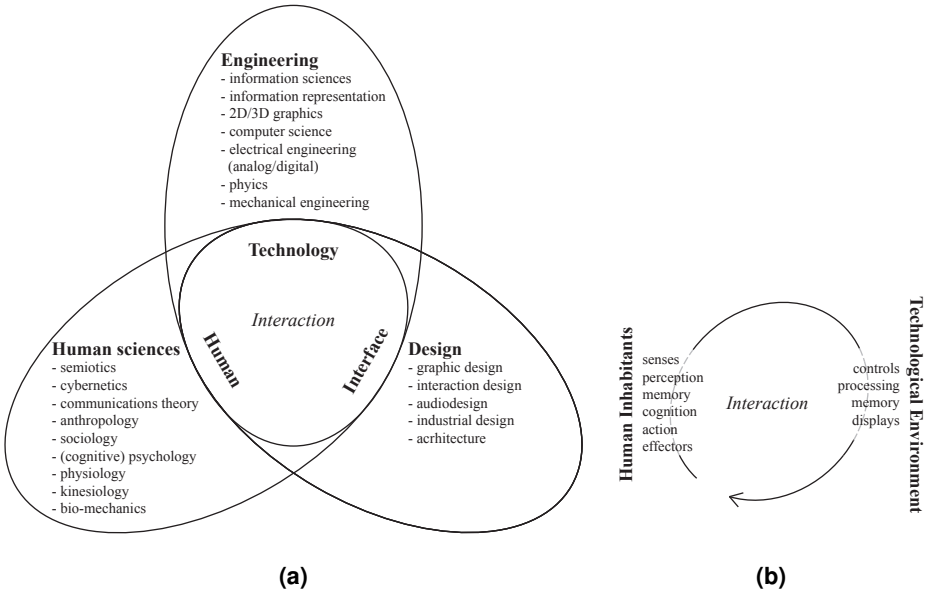


Figure 2.1: (a) The HCI domain emerges from multiple disciplines and (b) the interactions take place as a two-way process of control and feedback. Images adapted from Bongers [12].

Bongers [12] describes the interaction between human and computer as a two-way process of control and feedback, see Figure 2.1b. Effectors enable the human user to control the system, for example, through speech. The system takes this information in through its controls: input devices such as the sensors described in Section 2.4. The system then outputs a response through displays, for example, screens, loudspeakers and motors. These responses are perceived through the human senses after which the loop is closed. Sisson [192] describes a similar HCI loop that focuses more on the human that perceives, recognizes, comprehends, thinks about, formulates intentions, plans and performs actions that are based on, and that feed, the interaction. Another way to look at the two-way interaction between user and system is proposed by Norman [152] who describes the mismatch between our internal goals on the one hand, and, on the other hand, the expectations and the availability of information that specifies the state of the technological environment or artifact and how it might be changed [153]. Norman [152] names this the 'gulf of execution'. It describes the gap between the psychological language (or mental model) of the user's goals and the physical action-oriented language of the device controls via which it is operated. Likewise, the 'gulf of evaluation' is the difficulty of assessing the state of the system and how well the artifact supports the discovery and interpretation of that state [153]. In this thesis we focus on the hands gesturing in an intuitive way. We formulate 'intuitive gestures' as gestures that minimize the mismatch in Norman's 'gulf of execution'. In addition, we take a human perspective on the way that these gestures should take form. The hands form the effectors that

perform actions to control the system. The input devices or sensors that the system should employ to look at the user are based on the way that the effectors/hands gesture, not the other way around.

To help understand the interaction between human inhabitants and their technological environment we can define interaction levels in various ways. The human inhabitants of these interactive environments have goals that lead to tasks for them to perform with an interface. Bongers [12] speaks of tasks in terms of semantic, syntactic and lexical levels: the semantics describe the meaning of a message that is constructed out of lexical elements that are cast in a syntactic form. Nielsen [145] even describes alphabetical and physical levels below the lexical level, while Sisson [192] is content with only a physical level. In this thesis, the message or semantics is the direct manipulation of display contents in the form of interface tasks, for example, 'delete an object from the screen'. These tasks are executed in a syntax that is not so important in the work presented in this thesis. It can perhaps differ in terms of '<delete> <select object>' or '<select object> <delete>'. The focus of this thesis are the lexical elements that are used to implement the tasks '<delete>' and '<select object>'. Our aim is to explore the gestural representations, or simply gestures, that are suited to control large displays. These gestures are the physical components that make up the lexical elements in our HCI dialogues.

In this thesis we explore intuitive gesture-based interfaces. Our aim is to minimize the mismatch between the user's goals and the semantics, syntax and lexicon that the user needs in order to interact with the large display interface. In the remainder of this thesis we speak of goals that a user has that fulfills some intention. To complete her goal, the user formulates a plan of tasks that achieve subgoals, for example, navigate to point A, open item B, change contents C. Tasks are the semantics of the interaction. Commands are issued to execute a task. These commands are the lexical elements on which we focus in this thesis. The commands are given in the form of gestures.

2.2 The need for intuitive gestures

We now zoom in to focus on human-computer interfaces with large displays that are controlled through gesturing. One popular approach in HCI to implement these interfaces is the use of handhelds: one or more devices that the user holds in her hands and through which she can interact with the system [43]. It will not always be possible or desirable to use handheld devices for controlling these displays. We clarify this statement with three examples.

First, in a shopping mall, potential users will casually walk by an interactive large display while not having a handheld controller available [8]. It can be argued that mobile phones can perform such a role through using their increasingly sensitive, on-board cameras [70; 213], their abilities for data input through keyboard [129] or possibly novel interactions such as front and back-typing [231].

Second, in project-based teamwork, detailed results on the large display help structure and feed the discussion. The *e*-BioLab [177], see also Section 1.3.1, offers

large displays to do just that but it lacks a means for the discussants to control them directly. The hands might serve as a means to control the display by all discussants all the time. There is no need to hand over the controller.

Third, for entertainment purposes, the gaming industry is rapidly developing new means to include controller-free interaction [149]. Microsoft released a vision for future gaming experiences in June 2009 named 'Project Natal'[1] for the Xbox 360 game console. No controller is needed for a large variety of games that focus on full body pose recognition for input.

The common denominator for these controller-free large display interfaces is their focus on the hands for issuing commands [87]. A typical way to interact through gesturing is to introduce a gesture set that is designed to accommodate the sensors that are used. In that respect, gesture studies can often be reduced to pattern recognition [88]. Such approaches typically do not address *which* patterns should be recognized: the state of the art all too often imposes unnatural DOS-command-like gestures upon everyday human users. For example, Vogel and Balakrishnan [215] had a Vicon motion capture system available that they used to detect crude hand orientations and user distance to control an ambient display from various distances. A flat hand with the palm facing the screen meant 'open' while turning the palm to face the user meant 'close'. Such predefined, idiosyncratic gesture sets can be difficult for users to learn and use. Instead, Wexelblat [227] argues that natural gesture interaction is the only useful mode of interfacing with computer systems with the hands: "[...] one of the major points of gesture modes of operation is their naturalness. If you take away that advantage, it is hard to see why the user benefits from a gestural interface at all". Such natural interaction is based upon human behavior in everyday environments while performing everyday activities [22].

We discern two types of interfaces that observe and react to such natural human behavior [150]: pro-active and reactive. Pro-active interfaces look at and interpret user behavior in a largely implicit way so that the computer seems to disappear [198]. The interaction takes the form of a dialogue between two persons rather than between a human and a machine. The action-oriented language of the machine is translated to fit the user's psychological language of her goals. The interface contributes information in much the same way as a human discussant would, sharing it at appropriate moments. However, the user can only indirectly influence the information that is shared, and the moments when it is shared. It has been argued that multimodal and gesture interfaces extend beyond the traditional WIMP interface and that they should be pro-active [16; 158]. Pro-active interfaces can respond to the user, for example, to inform the user with calender information, when they are in close proximity to the system [214] or by switching music channels when they pick up a colored object [100]. There is no very strict line that distinguishes pro-active and reactive interfaces from one another. However, the tasks in pro-active interfaces depend heavily on the contents of the system and context in which they are performed, making it is impossible to identify elementary tasks for pro-active interfaces. Reactive interfaces, on the other hand, work in quite the opposite way

[1] `http://www.xbox.com/en-US/live/projectnatal/`, November 17th, 2009.

by focusing on more explicit commands [237]. In reactive interfaces, the language is more similar to the action-oriented machine language. Arguably, the user might feel more in control of the interface in explicit command-giving settings. However, we think that in both cases the user experience in terms of ease of use, learnability and enjoyability will benefit from gesturing that comes naturally [158].

In this thesis we explore the gestures that come naturally to our users and we consider those gestures to be intuitive. Referring back to Norman and Draper's 'gulf of execution', these intuitive gestures minimize the mismatch between the user's intentions and the action-command language that the system expects as input, see Section 2.1. We recognize that there may be very diverse causes why some gestures are considered to be intuitive by users while others are not. Users might, for example, rely on strong physical metaphors in their everyday lives and work, gestures that are typical from the culture that they grew up in [37], but it might also be that decades of indoctrination of mouse-based interfaces has created a new, technologically driven, metaphor that users consider as intuitive as well. In order to discover which gestures come naturally and to get a feeling why this is the case, we take a top-down approach in which the user has a central place.

2.3 Elementary interface tasks

Explicit command-giving is the basis for the type of large display interfaces that we focus on. Our aim in this thesis is to discover how existing large display interfaces might be controlled with the hands. However, it is not clear per se which elementary tasks lie at the heart of these interfaces. In this section, we define a set of elementary interface tasks for which we try to assign gestures in the remainder of this thesis.

Elementary tasks build up an interface by being repeated throughout the various facets of the whole interaction. The best known example of such an interface is the WIMP[2] design where point-and-click events are used and reused over and over again. By chunking together a series of low-level, elementary tasks, a whole interface can be constructed. Developers of a gesture interface should therefore focus on finding a set of elementary tasks with which they can construct an interface that is self-revealing, simple and flexible. Our focus here lies on how to control reactive interfaces that are operated through explicit command-giving. Although alternatives to the point-and-click paradigm have been explored, the point-and-click metaphor has a self-revealing nature, simplicity, and flexibility that is hard to beat [215]. It consists of moving the cursor (pointing) and then confirming that the target has been reached (clicking) [2].

The perspectives for the user and the system differ greatly with respect to which tasks are executed: consider Norman and Draper's gulfs of execution and evaluation [154], see Section 2.1. For the user, the interaction seems to be concentrating on navigating through, selecting and manipulating objects on a screen [13]. These tasks can consist of smaller subtasks that are chunked together sequentially [20], for example, navigation consists of one or more sequential point-and-click actions.

[2]Windows-Icons-Menus-Pointing.

For the system, these subtasks consist of one or more chunked actions that it detects through one or more sensors; typically a mouse with one or more buttons. For example, a chunked click action consists of depressing and releasing the left mouse button. The system perspective is tuned to the sensors that observe the user, for example, buttons, while the user perspective focuses on interacting with the data. Buxton [19] argues that, in order to describe the interaction in a more generalized way, human-computer interfaces should be described more from a human perspective. By doing so, the device, system or sensor that is used for input becomes less important. We now describe interactions with a gesture-based interface from both a system's and a user's perspective.

2.3.1 A system's perspective

Buxton [19] proposed a three-state model to represent the interactions such as point, select and drag for devices such as the mouse, see Figure 2.2. However, when looking at this model, we believe that it mainly describes the system states rather than how a user perceives the interaction. As an example of Buxton's model, a one-button mouse can be represented to be out-of-range (state #0) when the user is not touching it, tracking (state #1) when the user is moving it and dragging (state #2) when the user presses the button. Selection is done with a quick 1-2-1 state transition. The precise meaning of these three states varies (slightly) with the device or interaction technique that is being represented. For example, a stylus is out-of-range when it is lifted from its tablet, a joystick has no out-of-range state because it keeps tracking when untouched and a buttonless joystick does not have a selected state. A stylus can support two ways of clicking in this model: either using a button for a 1-2-1 state transition or by lifting the stylus for a 0-1-0-1-0 state transition. Additional sensors have been added in order to increase the sensing capabilities of devices such as the mouse. Hinckley *et al.* [80] built a touch-sensitive mouse (TouchMouse) that can sense when a user is touching it so that state #0 can be explicitly detected.

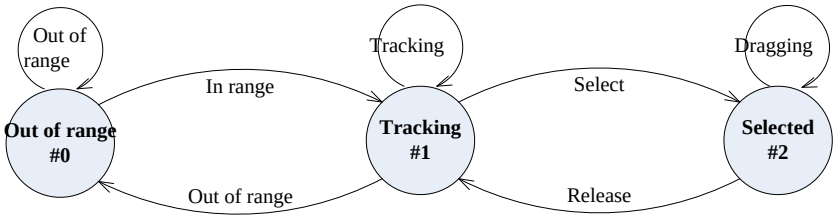


Figure 2.2: Buxton's three-state model for graphical input. Image adapted from [19], note that Buxton did not name his states in this manner.

Buxton's three-state model [19] can also describe input devices that work directly on the display surface. For such devices, for example, light pens and touch surfaces, a special case of the model applies with a direct transition between the #0 (passive tracking) and #2 (dragging) states because the system does not know what is being pointed before contact. Looking at the hands gesturing, Buxton's three-state

model describes such interfaces adequately but not fully. The user is out-of-range when not addressing the screen. By addressing the screen, the system switches to the track-state from which selection is possible. Manipulation of any (selected) contents is, however, not always a chunking of these three steps. For example, resizing a selected object [92], activating a selected object [2] or performing a 'right-mouse'-like action on a selected object [79], cannot be described.

In an effort to support a second motion sensing state on the PDA as with the mouse, a five-state model was proposed in which new states are added for each new mouse-button by Hinckley [79]. He added a hover-sate, for stylus or finger over the display, in addition to describing each button (i.e. left and right) with a select-state and a drag-state. However, when reviewing other devices such as the mouse, each button represents a specific type of interaction: selection (left-button) or manipulation (right-button). The approach by Hinckley would mean that for each new interaction two new states need to be introduced to the state model.

Chen [27] adds a fourth state to Buxton's three states. His fourth state can handle selecting an object followed by the user moving out-of-range while keeping the object selected. This four-state model cannot describe the way in which an object is manipulated, which, for all its flaws, was dealt with by the model of Hinckley [79], nor does it address tracking when the user moved out-of-range. Ahlstroem *et al.* [2] expanded Chen's framework with a fifth state that captures the way in which an object is manipulated. Ahlstroem *et al.* focused only on menu-selection tasks in interacting with cascading-menus. Their five-state model is capable of reaching all menu options with only vertical movements and short dwell times. However, due to their focus on menu-navigation, the five-state model deals mostly with when and where the button is released rather than describing added diversity in manipulating objects.

2.3.2 A user's perspective

The solutions we have seen so far have two main drawbacks in common: they cannot generalize over different manipulations in an interface and they describe the interface in system states rather than in terms of user goals despite Buxton's aim to take a more user-centered perspective. The different manipulation tasks can range from text input to resizing an object (e.g., window) to repositioning and orienting objects to deleting objects. However, if we were to describe all possible manipulation tasks the model would never be finished without a very strict context in which the model is to be applied. We feel that by adding one dynamic state to Buxton's three-state model, and by taking care to implement the resulting four-state model, manipulation tasks can be generically described by this model. In addition, to make the model more user-centered, we propose to remove the quick state transitions of Buxton's model that mainly facilitate pressing buttons and other actions that follow from a system's perspective. For example, the act of pointing is an elementary tasks that does not require, from the user's perspective, any subtasks. Navigation however, requires repeatedly performing point and select tasks [68].

Manipulations, such as resizing a window, are performed by actions no different

than other point-and-click events that can be modelled by Buxton's model. By adding a state that dynamically encompasses manipulation we gain a means to alter the contents in various ways. This is similar to Hinckley's approach where new states are introduced for each additional button [79]. It is important to note that for each context, for example, a presentation environment contrasting to a interactive art setting, in which this model is applied, different implementations of the manipulation state might be required. To prevent a rapid growth of additional states to this model, our fourth state, see Figure 2.3, is implemented in forms that are adapted to the context in which it is applied and the tasks that flow from this context. For each implementation of the model and the various manipulation states in it, each state transition might benefit from a different gesture. In a sense, this is no different from the variations that Buxton [19] introduced to model different input devices. So far, our four-state model can deal with one hand gesturing in out-of-range and in-range settings to select and manipulate display contents. Note that dragging (or repositioning) is, in our model, a form of manipulating.

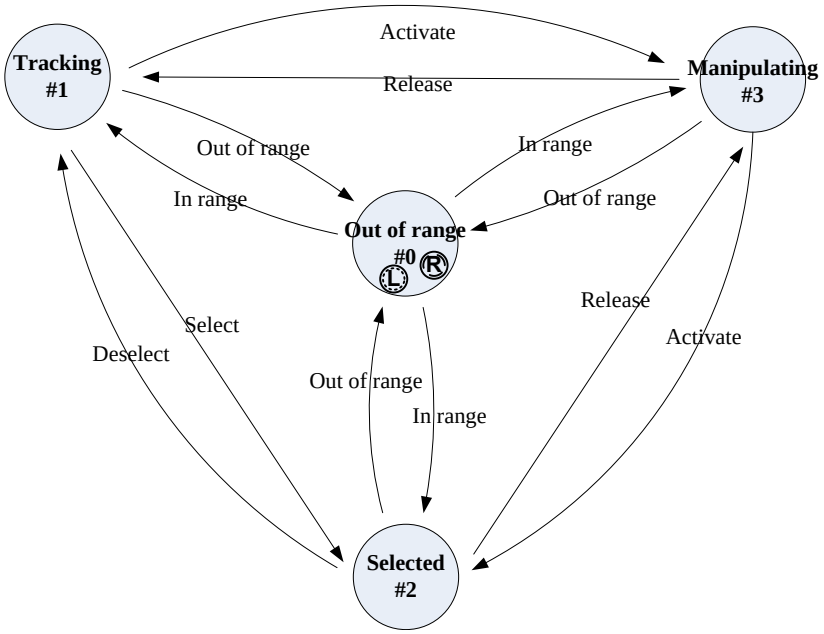


Figure 2.3: Our four-state model for direct free-hand interface input, extending the three-state model proposed by Buxton [19]. The manipulation state (state #3) can complete diverse tasks such as resize and activate. The two hands are represented by tokens (L and R here) that can move, separately or jointly, between states in a Petri net-style way.

In everyday life there are few tasks in which the second hand does not perform a role, if only by remaining passive [66]. Guiard's kinematic chain (KC) model describes the human motor model for two hands and it describes how the non-preferred hand (NP) serves as a reference frame for the preferred hand (P) [65]. The KC-model assumes two things. First, that the hands represent two motors that serve to create motion and, second, that these two motors cooperate with each other as if assembled in series, thereby forming a kinematic chain. Hinckley *et al.* [80] proposed to extend Buxton's three-state model to encompass two-handed input transactions

by using a Petri net representation. They introduced Petri nets as a solution for representing the inherent parallelism of two-handed input while preserving the flavor of the three-state model. In a user experiment to test the capabilities of this Petri net description, users held a puck in their NP-hand and a pressure-sensitive stylus in the P-hand [80]. The interface supported annotation, panning and zooming of maps without heavyweight mode switches, for example, having to click an icon in order to change the operating mode of the cursor. Hinckley *et al.* [80] argue that Buxton's three-state model treats their TouchMouse and a puck on a tablet very much the same while they argue that very different interactions are afforded. They extend the three-state model in two ways. First, they show how annotating states of the model with continuous properties while in that state is a useful parameter and can better describe the idiosyncrasies of various devices. Second, a distinction is drawn between out-of-range events based on touch versus those based on proximity of an input device to a sensor. Although Hinckley *et al.* claim that they restructured Buxton's model in a Petri net form, their model is in fact not a Petri net *pur sang*. In Petri nets, there are no edges that move between states, all state changes flow through conditional transitions [180]. However, even without such a strict Petri net implementation, the concept of tokens to represent the two hands is very useful.

Looking at our four-state interaction model in Figure 2.3, we can see that each hand is modelled with a token. The left-hand (L) and right-hand (R) tokens can transition between the four states separately or together. The way in which the tokens transition between states defines the specific unimanual or bimanual interaction, or the explicit commands that can be given in the interface. For example, when one token is in the selected state, the second token can manipulate (e.g., drag, resize, orient) the selected object in the manipulating state. Asynchronous bimanual gesturing is captured by allowing the two tokens to move separately from each other. Symmetric and asymmetric bimanual gesturing is captured by defining the transitions differently for each token, meaning, similar to implementing a two-button mouse versus a one-button joystick with Buxton's three-state model. In addition, for asymmetric bimanual manipulation tasks, we do not define the parameter changes that are represented by the movements of the hands [212]. Contrary to Buxton [19], our four-state model does not include the repeat edges at each state. In Buxton's three-state model those repeated-state transitions were included to describe remaining in a state, for example, to capture dragging in the selected state. In our view, this is mainly done to facilitate the system's point-of-view rather than the user's: the user is not doing anything new when, for example, dragging. By introducing the tokens to represent the hands, the tokens can simply remain in a state, thereby removing the need to include such repeated-state transitions. Note that we see each hand as an effector that can perform an action. It is possible to extend our four-state model to include tokens that represent individual fingers, or to abstract it so that one token represents both arms as Guiard's whole kinematic chain.

It is important to note that this model is not implemented in full for each interface that it describes. We feel confident that it can describe any gesture-based interface in which users give commands explicitly. Equally important, this model

describes the interaction from the user's point of view, not from that of the interface. As an example, consider the case where the user is dragging an image across two or more physically separated displays, as in the *e*-BioLab [178]. The physical space in between two displays is then not considered out-of-range as far as the user is concerned but it might be for the system's sensors.

2.4 Looking at gesturing

Technological developments enable new forms of gesture-based interactions. Here, a state-of-the-art overview of enabling technologies is provided. We omit an overview of display technologies that provide feedback to the user here and redirect the reader to Fikkert *et al.* [43, Sec.3.2] for an extensive overview. We describe both commercial products and research concepts. The focus in this section is on enabling technologies that 'look at' the hands and arms gesturing in the hands, at arm's length and at distal scales. We distinguish four enabling technologies: handheld devices, haptics, vision and wearable sensors. Please note that we strive to give the reader a feeling of the possibilities in looking at gesturing that is offered by various types of sensors. This overview is by no means exhaustive.

2.4.1 Handheld devices

A pragmatic approach to implement a gesture-based interface is to use handheld devices with sensors that are tightly mapped to the interface they control. In many of these devices, buttons are added to enable additional interactions that, for example, mimic mouse-buttons. Handhelds explicitly signal that someone is in control or not which makes them useful in multi-user settings [138]. The interface responds predictably when the device is given to another user or put on a table. One issue that prevents these devices from being readily used in public environments is that they are not always available there. We believe that the mobile phone offers great potential for this [61; 213]. Handheld devices are reported to a large extent to induce significantly more stress than desktop-based devices [7].

Ergonomics influence the design of handheld devices [164]. Using a precise pinch grip in the dominant hand, and a more crude grip in the non-dominant hand, Stefani and Rauschenbach [196] designed two distinct handheld devices with which CAD drawings could be manipulated, see Figure 2.4a. Tangible objects that represent their virtual counterparts offer an alternative device design [56; 63; 81]

Pointing towards the screen and pointing at objects depicted on it can be done by either tracking the device externally or internally. External tracking via computer vision solutions is described in Section 2.4.3 and via (absolute) position sensors in Section 2.4.4. Internal tracking is done in the popular Nintendo Wiimote by detecting fixed infrared light sources with an on-board camera, see Figure 2.4b, or by using a laser pointer that projects onto a surface [113]. Relative position sensors are also applied in handheld devices. A myriad of 3D mice, which exploit accelerometers, that detect motion, and gyroscopes, that detect rotation, have been

released to the market, for example, see Figure 2.4c. The most popular example of the use of accelerometers is the Nintendo Wiimote, see Figure 2.4b.

(a) (b) (c)

Figure 2.4: Handheld devices for arm's length and distal scale interactions. (a) The 'bug' and 'dragon fly' [196]; (b) The Nintendo Wiimote; (c) Several 3D mouse implementations[1].

2.4.2 Haptics

In the absence of handheld controllers, interacting with a display at arm's length and distal scales suffers from the lack of haptic feedback in situations which precision and feedback are crucial [41; 42]. Here we introduce some solutions that add haptics to such interfaces. The human haptic sense is composed of the kinesthetic sense and the tactile sense. The kinesthetic sense detects force and motion effected by the human body. The kinesthetic sense is a means of input for the system. Touch is sensed by the user through the tactile sense. It is produced as one means of output from the system. The tactile stimulus is affected by different kinds of receptors under the human skin which varies along the human body, for example, at the fingertips one can sense distances of one millimeter [71]. Allowing a user to touch a virtual object provides additional insight. Application areas for haptics include endoscopic surgery [116], sculpting [109], CAD drawings [103] and the disposal of explosives [121].

(a) (b) (c) (d)

Figure 2.5: Examples of haptic devices for arm's length and distal scale interactions. Immersion's[2] (a) CyberTouch and (b) CyberGrasp; (c) The RM II Hand Master [78]; (d) The DiamondTouch multi-touch surface [44].

[1] `http://www.3dconnexion.com`, September 8th, 2009.
[2] `http://www.immersion.com`, September 9th, 2009.

Kinesthetic feedback can be generated by an exoskeleton equipped with actuators, see Figure 2.5b. Such accurate kinesthetic feedback can be used for rehabilitation [78], see Figure 2.5c. A less precise solution uses only actuators on various places on the hand, see Figure 2.5a.

At arm's length, multi-touch sensitive displays have taken flight since Han [73] introduced cheap, off-the-shelf multi-touch technology. Touch-sensitive surfaces generate an implicit tactile response when users touch the surface. It provides a moment of rest for the hands while also being a strong indicator of when a response from the system might be expected. Fikkert *et al.* [44] provide an overview of various multi-touch solutions such as capacitive coupling [35] (see Figure 2.5d), optical recognition [98; 131] and stereo vision [228]. To provide kinesthetic feedback on touch-sensitive surfaces, Wagner *et al.* [220] used RC servomotors to distort the surface.

2.4.3 Vision

Robust, unobtrusive detection of finger, hand and arm gestures is one holy grail in HCI. Most solutions focus on applying computer vision solutions in the monochrome, infrared or visual spectrum. We keep this topic brief as Gavrila [57], Moeslund *et al.* [136], Poppe [168] and Jaimes and Sebe [93] already provide excellent and highly detailed surveys of this field. Apart from being unobtrusive to the user, a big advantage of optical tracking is the low latency time and the immunity to environmental factors, for example, ferromagnetic materials in the case of magnetic tracking. On the downside, optical tracking is highly sensitive to occlusion [124]—objects disappearing in the camera image behind other objects—that can take the form of self-occlusion, occlusion by the environment and occlusion by other users. In most cases, solutions that employ multiple cameras are used to overcome the occlusion issue [235]. In addition, various filter techniques, for example, Kalman [224], can be employed to predict where a target is located when line-of-sight is lost.

Markerless [137] and marker-based [183] tracking are the two major categories into which most vision-based gesture interfaces fall. In markerless approaches, the main challenge is to detect and track the fingers and hands robustly. Issues such as limited resolution, (self-)occlusion and cluttered background add to this challenge. Marker-based tracking typically focuses on active [163] or passive markers[3] in various sizes, shapes and patterns on fingers [104], hands [214] or handheld objects [84]. One solution to increase contrast in camera images is to use coloured gloves [89]. Laser-tracking has been used to track finger tips in midair [24].

2.4.4 Wearable sensors

One of the most pragmatic yet invasive methods to look at the hands gesturing is to put sensors directly on them. We have already seen haptic solutions in Section

[3] `http://www.vicon.com`, September 9th, 2009.

2.4.2, here we describe the various types of sensors and how they are used. Motion tracking suits offer complete capture of bodily movements, see Figure 2.6a. On a smaller scale, data gloves have been used for decades to detect the high-resolution movements of the hands and fingers, see Figures 2.6c and (b). These suits and gloves typically differ in the number and types of sensors, sensor resolution and sampling rate. All wearable sensor solutions are sensitive to the precision in which the sensors are placed and how well they stay in place. Gloves and suits do not fit every person in the same way, so that sensors may be misaligned.

Figure 2.6: Examples of wearable sensors for arm's length and distal scale interactions. (a) Animazoo's[4] Gyspy 6; (b) XSens'[5] MVN; (c) The Fakespace[6] Pinch glove; (d) The Inition[7] P5 glove.

These suits and gloves can be tracked with relative (inertial) and absolute (magnetic, optical) sensors. The hand shape in a dataglove is typically detected with bend sensors that can detect the relative flexing of, for example, the phalanges of each finger or the jaw of the wrist. Inertial sensors measure relative changes in their location and as a result, the location that they detect will drift from the calibrated position over time. Magnetic sensors, such as Ascension's Flock of Birds[8], detect the location and orientation in a calibrated space, making them more robust to drifting. However, these sensors can be influenced or distorted by ferromagnetic metals and copper [151]. The principles of optical tracking solutions do not differ from those described in Section 2.4.3. Similar to magnetic tracking, ultrasonic tracking can detect the position and orientation of objects in a bounded area [234]. In terms of physics, ultrasonic tracking is similar to optical tracking.

2.5 Summary

We have shown that the human-computer interaction research field is a meeting place of multiple disciplines. The HCI field studies the relationship between hu-

[4] `http://www.animazoo.com`, September 9th, 2009.
[5] `http://www.xsens.com`, September 9th, 2009.
[6] `http://www.fakespace.com`, September 9th, 2009.
[7] `http://www.inition.co.uk`, September 9th, 2009.
[8] `http://www.ascension-tech.com`, September 9th, 2009.

mans and their technological environment. The interaction between humans and the technological surroundings is a two-way process of control and feedback. The human user has intentions that she tries to formulate as actions while the action-command language that the system expects as input might be entirely different. HCI research tries to minimize this mismatch by building intuitive interfaces.

Gesture-based interfaces are the intuitive interfaces that constitute the focus of this thesis. We described the need for these intuitive gesture-based interfaces and where they might be employed in existing technological environments. These interfaces communicate a message that has a syntax and a lexicon. We have defined the elementary tasks that are present in a gesture-based interface for the user to perform. These tasks are placed in a four-state model: out-of-range, tracking, selected and manipulating. Tasks use commands to switch between these states. The commands are given in the form of gestures made with one or both hands.

This chapter was concluded with an overview of technologies that are suited to automatically detect, track, recognize and interpret humans gesturing. These handheld devices, haptics, vision and wearable sensor systems provide the eyes and ears of the environment that the user inhabits. In Chapter 3, we will now describe what gestures are, how they can be categorized and how gestures have been used in HCI.

Chapter 3

Gestures

> "A gesture cannot be regarded as the expression of an individual, as his creation (because no individual is capable of creating a fully original gesture, belonging to nobody else), nor can it even be regarded as that person's instrument; on the contrary, it is gestures that use us as their instruments, as their bearers and incarnations."
>
> **Milan Kundera**
> Czech writer, born 1929 – Immortality, 1999

The focus of this dissertation lies on command-giving through gesturing to large displays that are placed beyond arm's length. The previous chapter described the context in which gestures will be applied: human-computer interaction. In this chapter, we clarify what is meant by the term gesture in Section 3.1. Second, we describe the characteristics of both communicative and manipulative gesture classes in Section 3.2. These classifications provide a useful tool to understand why gestures are used or avoided in human-computer interaction. Third, Section 3.3 describes the process of detecting and recognizing gestures in a human-computer interface. Fourth, Section 3.4 describes how gesture sets have been built and used in HCI studies, in commercial interfaces and which gestures have been observed in the *e*-BioLab, see Section 1.3.1. We summarize this chapter in Section 3.5.

3.1 Definition

The term gesture is defined and used in diverse ways depending on the context in which it is used. Gesture theory used in human-computer interaction (HCI) originates from linguistics, anthropology, cognitive science and psychology [102]. As such, 'gesture' has been translated in HCI as a motion of the hands [14; 41; 143], facial expressions [127], gaze tracking [130; 232], head movements [239], hand postures [88; 146] and whole body postures [188]. Although it has many potential meanings, McNeill [134, p.1] and Kendon [108, pp.1–2], like many other researchers, denote the term gesture as communicative hand and arm movements that occur during spoken human dialogue. Interfaces that converse with their users in

a human-like way often target these communicative gestures as one of many input modalities [16]. In contrast, manipulative gestures—self and object touches—do not communicate meaning but they have proven to be an intuitive means of interacting with a system, examples of which are, tangibles [165; 186] and multi-touch interfaces [44]. It should be noted that in gesture research that studies the relation between speech and gesture, manipulative gestures are not considered to be gestures at all. Given the nature of that research this is to be expected [106; 134].

These two gesture types, communicative and manipulative, can be nicely placed on "Kendon's continuum" of gestures [108, pp.104–105]. Gestures are, at one end of this continuum, used in conjunction with speech insofar that users are only marginally aware of their gesturing. These movements of the hands and arms during speech are idiosyncratic and spontaneous [106]. At the other end of the continuum, gestures are used independently from speech. There, gestures are made fully consciously and they are compositional and lexical in structure. Due to the focus of this dissertation on gesture-based interfaces in HCI we include both communicative and manipulative hand and arm movements in our interpretation of the term gesture.

McNeill [134] states that gestures are global, synthetic and never hierarchical. Contrary to speech, gestures do not form a whole (sentence) based upon individual parts (words). Gesticulation is global and synthetic: "The meanings of the parts of a gesture dare determined by the whole, and different meaning segments are synthesized into a single gesture" [134, p.41]. McNeill claims that gestures are noncombinatoric in the sense that two gestures produced together do not combine to form a larger, more complex gesture. This notion is also found in HCI where gestures are not considered hierarchical but sequential [59]. As we will discuss in Section 3.3, the gesture recognition process starts with (automatically [96; 122]) segmenting continuous gesturing before it enters the recognition and interpretation phase [93; 136].

With respect to the generation of gestures in the human mind, we mostly find ourselves trying to determine the relationship between speech and gestures. It is commonly accepted that, although the one can be produced without the other, gestures and speech are closely related [18; 134]. However, there is controversy as to how gestures are produced in this relationship. Some research indicates that gesture representations are part of the speech that is being produced [18; 120]. It has, for example, been shown that spatial gestures are more common when the accompanying speech content is also spatial [175].

3.2 Gesture types

To gain insight into the gesture taxonomies that are used in the field, we categorize gestures from two angles. First, in Section 3.2.1, we describe the more traditional point of view that focuses on gesturing in the context of human-human communication. Literature on linguistics, anthropology, cognitive science and psychology typically categorize gestures as such. Second, in Section 3.2.2, we look at gestures from a HCI point of view. In pro-active interfaces, gestures are interpreted by the

system as part of a dialogue in which speech is a main contributor. In reactive interfaces, gestures are used more explicitly for issuing commands which can be stand apart from speech. After introducing these taxonomies, we briefly describe the similarities and differences between them. This section concludes with a summary of the gesture classes that are used in the experiments reported in Part II of this thesis.

3.2.1 The traditional taxonomy: human communication

Well-known gesture classification schemes are those by Efron [37], Ekman and Friesen [39], Freedman and Hoffman [54], McNeill [134] and Kendon [105]. It should be mentioned that Efron's work is the shared forefather of the other classifications. As a result, there are many similarities in the classes in each of these schemes and often gesture classes differ only in name. For example, McNeill's beats are the same as Ekman and Friesen's batons. We follow Kipp [110] in his overview of six gesture classes that flow from these four classification schemes because those classes are well-documented, recognizable and they cover all possible human gestures. Kipp divided gestures in non-communicative and communicative gestures. Adaptors form the sole class of non-communicative gestures while iconics, metaphorics, beats, emblems and deictics are communicative gestures, see Figure 3.1. Please note that these gestures are placed on the so-called "Kendon's continuum". This continuum serves to illustrate that it is not always possible to discretely categorize gestures. For example, iconics and metaphorics are separated by a shady grey area [110, p.36].

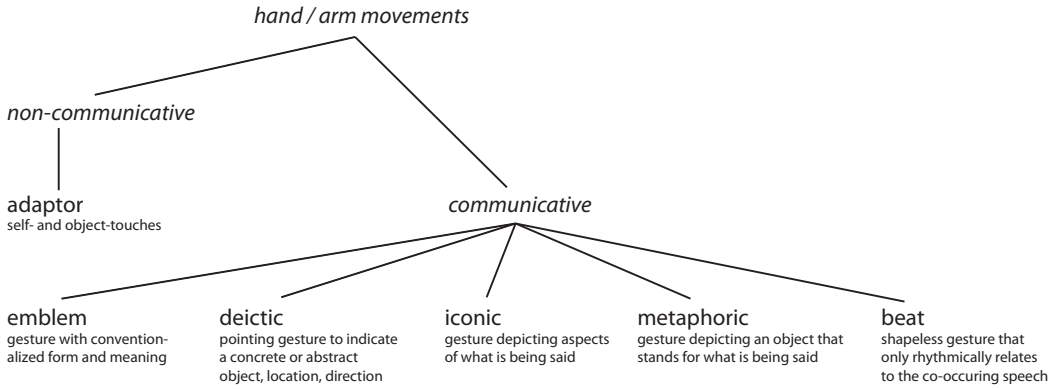


Figure 3.1: Kipp's more traditional taxonomy, reproduced from [110, p.32].

McNeill [134, pp.86-104] defines a gesture space as a shallow disk in front of the speaker, see Figure 3.2. The gesture space is a spatial division into regions that has been used to study gesturing. For example, iconics fill the center-center space, metaphorics congregate below in the lower-center space, and deictics extend to the periphery. McNeill's gesture space is a useful tool for describing how gestures utilize space in different ways which, McNeill argues, is a strong justification for subdividing gestures according to the six gesture classes: iconics, metaphorics,

beats, emblems, deictics and adaptors [134, p.88].

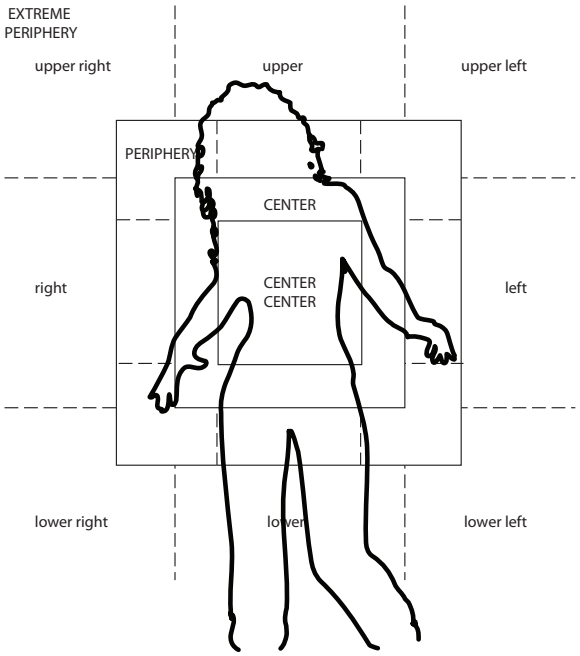


Figure 3.2: McNeill's gesture space, image adapted from [134, p.89].

Iconics

Iconic gestures illustrate what is being said and as such they bear a close formal relationship to the semantic content of speech [134, p.12]. They do so by depicting some property of the speech referent. It is rare that iconics have a standard form, they are often made up on the fly [110, p.35]. Ekman and Friesen [39] subdivided iconics, or 'illustrators' as they were named, into six subclasses. Three of those, spatial movements, kinetographs and pictographs, unite to form the class of iconic gestures as we use it here. Spatial movements describe, for example, the spatial constellation of the team players in an underwater hockey match. Iconic gestures resemble metaphorics but instead they illustrate a path or object shape. Cassell *et al.* [23] argued that iconics also communicate the viewpoint from which the action is narrated. For example, a speaker might take the viewpoint of Granny hitting Sylvester with an umbrella when he tries to eat Tweety, or Sylvester *being* hit with an umbrella. Sowa and Wachsmuth [195] used iconics to refer to virtual objects that are depicted on a large display, by describing the object shapes through gesturing.

Metaphorics

Like iconics, metaphorics are pictorial. The pictorial content indirectly represents an abstract idea rather than a concrete object or event [134, p.14]. A metaphoric

gesture presents a metaphor as a bounded, supportable, spatially localizable physical object. McNeill [134, p.146] called the gesture a 'sign', the imaginary object the 'base' and the concept of, for example, a story, the 'referent'. Iconics are rather similar to metaphorics and if base and referent are identical or very similar, the gesture is iconic [110, p.36]. An example of metaphoric versus iconic gesturing is the utterance "this match was really hard because of their very experienced defender". The speaker first holds out one hand to represent the match: an abstract idea (metaphoric gesture). Secondly, he uses his other hand to represent the player: a concrete object (iconic gesture).

Beats

Rhythmic hand movements that accompany speech but where the hand shape bears no relation to the speech content are categorized as beats. These gestures often look like beating to, for example, music; the hand moves along with the rhythmical pulsation of speech. Beats have just two movement phases: in/out, up/down, etcetera. Beats also highlight discontinuities in the temporal sequence. An example of beats is the team coach reflecting on the match beating his hand three times in midair while saying "we have to improve attack, defence and endurance". By gesturing in this manner, the speaker is highlighting important sections of the spoken dialogue.

Emblems

Emblems are gestures with conventionalized form and meaning so that they can be expressed even in the absence of speech [110, p.33]. Emblems are also considered to be conventionalized in both their form and meaning, for example, 'thumbs up'. This makes emblems culture-specific [37]. However, these conventions may vary between subcultures. For example, scuba divers use 'thumbs up' for signalling the ascent to the surface. Similarly, Gullberg *et al.* [69] showed that native speakers of Dutch and French gesture differently when describing how to put, for example, a cup on a table, while the act of putting the cup on the table did not vary. Emblematic gestures can directly be translated to words that are sometimes uttered in conjunction with the emblem, for example, uttering "Sylvester was spying on Tweety using binoculars" while cupping the hands in the shape of that object.Emblems are especially used when the verbal channel is somehow restricted, consider examples such as scuba divers, crane operators and aircraft marshals.

Deictics

Pointing in narrative is known as a deixis. The function of deictics is to indicate objects and events in the concrete world but they also play a part even where there is nothing objectively present to point at. These gestures are "pointing movements whose function is to indicate a concrete person, object, location, direction but also to point to unseen, abstract or imaginary things" [120]. In most gesture taxonomies, deictics are found and they belong to the best studied of gesture types, especially in

HCI settings. Kipp [110] divides deictics into concrete (object) and abstract pointing (imaginary) which is a useful distinction when looking at human-computer interfaces. Concrete pointing is likely in human-computer interaction as it is a straightforward implementation of real-world object manipulation in addition to the way in which mouse interfaces in the WIMP paradigm are implemented. Abstract pointing can be added only on a semantic level due to its need for context, for example, a user can be standing with his back towards a display talking to someone while referring to something on the display by pointing over the shoulder with her thumb. Kendon [108] specifies pointing gestures into no less than seven subclasses which are mostly variations on hand shape rather than that they vary in any semantic way.

Adaptors

Adaptors are movements that are not considered part of communication by a recipient [39]. In conversations, self and object touches that are not considered part of the communication, for example, scratching one's ear lobe [110, p.32], are dubbed adaptors. In most gesture taxonomies, adaptors are not considered as gestures. For example, McNeill [134, p.78] even excludes adaptors from his definition of gestures. Kendon [108, p.97] notes that there are three types of adaptor that, according to their functions, can be distinguished. Self-adaptors satisfy self and bodily needs or they perform bodily actions such as scratching the head. Alter-adaptors manage emotions and maintain prototypic interpersonal contacts. These adaptors might be considered communicative in nature as well. For example, folding the arms as to protect oneself also non-verbally communicates that the speaker or listener feels threatened. Object-adaptors are used in tasks that involve an object such as smoking or tapping a pencil.

3.2.2 A gesture taxonomy for human-computer interaction

HCI researchers who build social actors that communicate with humans in the way humans do support the traditional viewpoint on gestures [179]. Although they have their use, as described in Chapter 1, these interfaces are beyond the scope of this thesis. Karam and Schraefel [102] followed Quek *et al.* [174] in defining five gesture classes that focus on the process of interacting instead of highlighting human communication. The five gesture classes that they define—gesticulation, manipulations, semaphores, deictics and language gestures; see Figure 3.3—nicely encompass the tasks that one would wish to accomplish with a computer interface [13]. Clearly, other similar taxonomies exist, for example, the work by Pavlovic *et al.* [166], but, because their taxonomy generically describes human-computer interactions through the hands, we will focus on the one proposed by Karam and Schraefel [102]. Deictic gestures in this taxonomy are identical to those described above in Section 3.2.1 which is why we omit its description here. We do include a brief overview of the use of deictics in HCI. It should be mentioned that Karam and Schraefel also introduced gesture taxonomies based on application domains,

enabling technologies and system responses that help to describe human-computer interactions.

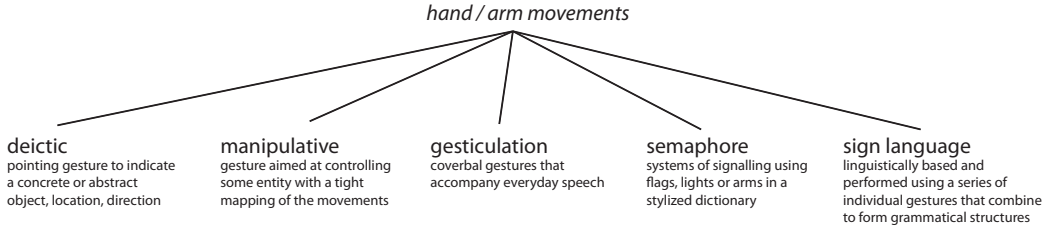


Figure 3.3: HCI gesture taxonomy by Karam and Schraefel, reproduced from [102].

Deictics

Deictics are often the central topic of research in gesture-based interaction with ambient intelligence [147]. Deixis has been used to identify objects in virtual reality applications [242], pointing out objects to others in CSCW applications [17], for targeting appliances in ubiquitous computing or robotics [144], for desktop applications [225], communication support [118] and remote collaboration [112]. Deictic gestures are considered by some as the basis for communicating with machines as equal partners in communication [16]. This idea is often described as the holy grail of human-computer interaction. In his classical work, "put-that-there", Bolt [11] already showed the intuitiveness and potential of deictic interfaces. Users pointed to a target (location) and issued commands via speech to either manipulate or select an object.

Manipulations

Quek *et al.* [174] defined manipulative gestures as tightly mapping the movements of the hand/arm to the movements of some virtual object in the interface. The traditional mouse interface does this in an indirect manner by translating the movements of the mouse to the cursor with some adjustments for speed. Karam and Schraefel [102] distinguish between gesturing in midair and on some surface but do not differentiate between direct and indirect mappings [191]. Another form of manipulative gestures uses tangible objects, for example, a doll's head that represents a MRI scan [81], to control an interface. Even though Karam and Schraefel do not describe methods to interact with 3D visualizations other than using tangible objects, other direct [7] and indirect [211] mappings have been proposed. A last form of manipulative gesturing is used to control real-world physical objects such as robotic arms [50; 90] and wheelchairs [239].

Semaphores

Semaphores are systems of signalling using flags, lights or arms. Quek *et al.* [174] extended this definition so that a semaphore is "any gesturing system that employs a

stylized dictionary of static or dynamic hand or arm gestures". Semaphoric gestures do not form the majority of signs or signals that communicate information in human interactions even insofar that they are, by some, not considered intuitive gestures [227]. However, they are a pragmatic solution for enabling distance computing in smart environments [149] and to reduce distraction from a primary task [101]. Unlike manipulative gestures which are mainly dynamic, semaphores can involve static poses as well as dynamic movements. Such semaphoric gestures can be performed with the fingers [64], hands [181; 228], head [239], arms [11], feet [46], body [168] or even hand-held objects. Strokes or marks that are made with a mouse [3], stylus [82], hands or fingers [186] are also considered to be semaphores.

Gesticulation

Gesticulations, or 'coverbal gestures' [142], are regarded in the literature as one of the most natural forms of gesturing because they are an integral part of human dialogues [106]. As a result, gesticulations are commonly used in combination with conversational speech interfaces [115; 167; 226]. Gesticulations are idiosyncratic, spontaneous movements of the hands and arms during speech and do not require the user to perform any poses or to learn any gestures other than those that naturally accompany everyday speech [102]. Thus, they are unlike semaphores, which are pre-recorded or trained for automated recognition, and unlike manipulations that tightly map movements to the interface.

Language gestures

Gestures used for sign languages are often considered independently of other gesture styles. Based on linguistics, sign languages sequentially combine individual signs that form grammatical structures [155]. These signs consist of basic components such as hand shape, orientation, location and movement. Sign gestures are not all symbolic, some are mimetic (pantomimes) or deictic, although most gestures still remain in the symbolic gesture class [134]. In sign languages, other modalities such as facial expressions and body posture are very important for sign meaning [202]. Because sign languages are lexically and grammatically complete, they are often compared to speech with respect to the processing that is required for their recognition. HCI applications that target these gestures focus on communication, for example, teaching sign language to children [126].

3.2.3 Overlap between the taxonomies

So far, we have seen two points of view that describe human gesturing. The traditional view categorizes gestures as they occur in human dialogue. This entails that speech is a large contributor to the occurrence of gesturing. The HCI point of view looks at interactions in order to control computer systems rather than analyzing gestures that accompany a spoken dialogue per se. So what are the differences between these points of view? The main differences flow from the application areas:

analyzing gestures that accompany speech in human dialogue and gestures for interacting with computerized systems. We compare the five categories of the HCI taxonomy by Karam and Schraefel with those summarized by Kipp. Please note that these mappings are not perfect and that there is overlap.

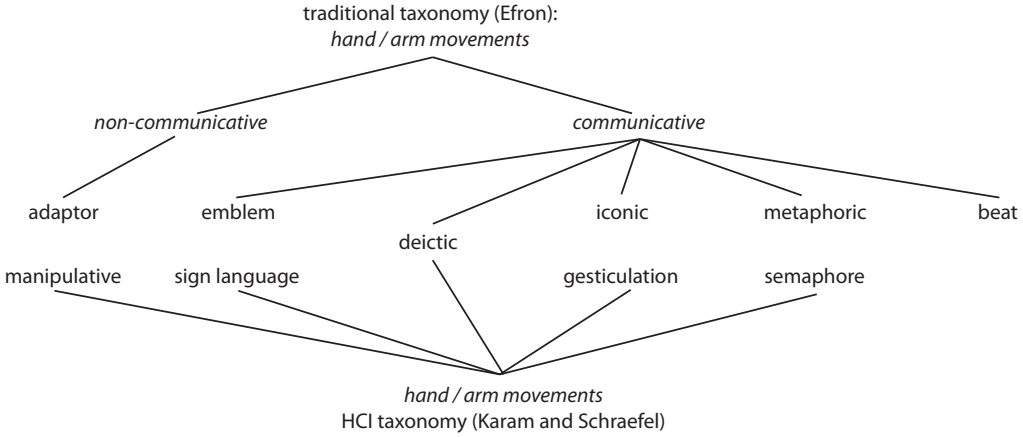


Figure 3.4: Overlap between Karam and Schraefel's HCI and Efron's traditional taxonomies. Note that there are *no* direct couplings between the gesture classes other than the deictics class. Classes are spatially positioned so that the best matching classes in the other taxonomy are nearby.

The overlap between the two taxonomies is depicted in Figure 3.4. Deictic gesturing is the same in both taxonomies. Manipulative gesturing controls some entity that is tightly mapped to the interface. Although adaptors do not communicate meaning, manipulative gestures can only be considered to be adaptors. Gesticulations are spontaneous gestures that a signer does not have to learn. Iconic and metaphoric gestures closely match such gestures. However, semaphores also adhere to this definition so there is an overlap. Sign languages are conventionalized in their form and meaning and they match best to emblematic gestures. Signed languages are not part of deictic gesturing because they are based to a great extent on co-occurrence with facial expressions, changes in body posture and even speaking the words that are being signed.

3.2.4 Gestures in this work

The above two gesture taxonomies are useful tools to embed this work in the existing HCI literature. We do not believe that either one of these taxonomies is best suited for our work as both tend to comprehensively describe human gesturing from different angles. We only focus on parts that we will describe now. Looking from the point of view of the traditional gesture taxonomies of McNeill and Kendon, the focus in this work lies on adaptors and emblems. Adaptors stand apart from the communication because objects or the self is manipulated while emblems can communicate information in the absence of speech through conventionalized form and meaning. With respect to the taxonomy of Karam and Schraefel, we focus on manipulative and semaphoric gestures in this work. Manipulations, like adaptors, see the

user grabbing and moving or touching an object while semaphores combine stylized static poses with dynamic movements. The common denominator here is that the gestures manipulate objects or self and that they do not require information that is communicated through speech.

3.3 Gesture recognition process

Before a computer system can translate hand movement and shape into meaning it needs to recognize a gesture as such. Gesture recognition typically consists of three steps [218]. First, gesture segmentation automatically [96] or manually cuts a sequence of movements into pieces that define where gestures start and end [122; 194]. Kipp [110], among others [173], defines this as gesture phrases. Second, each gesture phrase is classified in an (often predefined) gesture class in the system [41]. Third, the gesture phrase is directly mapped to a system response, mapped to a model of the hand for further interpretation [238], or it is correlated with co-occurring speech [111]. Recognizing gestures via segmentation can, in this light, be reduced to a pure pattern recognition problem [87]. Typically, the segmentation process focuses on features such as minima in hand velocity or differential in finger flexure [184], large changes in motion trajectory angles [212] and ignores large finger movements. Other solutions target training neural networks, hidden Markov models (HMMs) [144], applying principle component analysis (PCA) or other pattern classification techniques [155]. We will not go into further detail on solutions for these challenges.

Gesture segmentation produces gesture phrases. These phrases are built up out of three phases: preparation, stroke, retraction [38; 108; 134; 218]. The preparation phase consists of a movement that sets the hand in motion from some resting position, for example, placing the hands below the waistline [195]. The stroke phase, also known as nucleus, contains the most explicit hand movements, or, better said, "[the stroke] is the phase of the excursion in which the movement dynamics of 'effort' and 'shape' are manifested with greatest clarity" [108, p.112]. In many applications, it is the stroke that provides the most information [110; 195]. Retraction, which is also known as recovery [108], then moves the hand back to a rest position. The boundaries between phases are subjective and sequence-dependent, which results in a diverse range of segmentation solutions [96]. In addition, the boundary between retraction of the previous gesture and preparation of the next gesture can be fuzzy ([110]) or even non-existent when two gestures closely follow each other.

3.4 Defining gesture sets

It has been argued that there is no gesture dictionary because gestures do not map on a one-on-one basis between meaning and form in daily gesturing between humans [22]. However, conventionalized gestures, for example, emblematic gestures,

in human communication have to be learned by the signer yet they are widespread in both local and global social groups. In fact, popular gestures, for example, gang signs, travel the globe with relative ease by being absorbed into social groups. We therefore think that it is viable to construct a gesture set for interacting with large display interfaces.

Requirements for selecting the gestures in a gesture interface are formulated by Cohen [30]. He argues that the gestures should fit a useful environment, that the system can recognize non-perfect gestures, that the system can interpret both a gesture's static and dynamic information components and that the gesture is recognized as quickly as possible, even before the full gesture is completed. Nielsen *et al.* [146] came up with similar requirements but more from a user perspective. They required gestures to be easy to perform and remember, intuitive, metaphorically and iconically logical towards functionality and ergonomic; not physically stressing when used often. However, in most (experimental) gesture interfaces, an idiosyncratic gesture set is defined for a limited set of tasks [132; 215]. Moreover, the gesture that is selected is often more technology driven than user driven. The sensor in the interface determines the 'best' gesture for a task, for example, Agarawala and Balakrishnan [1] describe BumpTop in which the shape of complex cursor-trajectories have to be learned. In this section, we describe which gestures or gesture sets have been proposed in both literature and commercial products for explicit command giving.

No way to 'click'

By addressing purely gesture-based input we come across the problem of how to select or manipulate objects in the absence of other modalities that can 'click'. A popular solution is to use dwell time thresholds that activate a select command whenever the user points to a target for some time, with a hand-held device [113], extended index finger [214] or eye-tracker for gaze location estimation [241]. Even though this solution is a simple one, it introduces a fixed, constant lag insofar that the interactions can suffer from the "Midas touch effect" [74][1]. With only the hand, we also need to consider that depressing a physical button or tapping a display surface produces a kinaesthetic feedback that confirms the click action. When beyond arm's length, there is no such surface to touch which will degrade the performance significantly when manipulating virtual objects [223]. Vogel and Balakrishnan [215] argue that the hand itself can serve as a source of kinaesthetic feedback that confirms gesture-actions through some tension in the hand. Grossman *et al.* [64] designed a gesture for clicking named *ThumbTrigger* in which the hand is shaped like a pistol: the thumb and index finger are extended while the rest of the hand is closed in a fist. With *ThumbTrigger*, clicking was done by pressing the thumb on the (bent) middle finger as if pressing an invisible button. Vogel and Balakrishnan [215] argue that a

[1] King Midas wished that he could turn everything he touched into gold. Of course, after having his wish granted, his food and drinks now turned into gold as well so that he had to beg to be delivered from starvation. In HCI terms, dwelling on an object to select it is a common practice. When the dwell-time is too brief, every becomes selected, without the user intending to do so.

click or clutch action should be designed to minimize hand movement side effects that will influence pointing precision.

3.4.1 Experimental gesture interfaces

Finding suitable gestures for tasks can be done from either a human perspective or from a system perspective. When a gesture set is designed from the human perspective, the gestures are categorized by terms such as intuitiveness, naturalness and ease of use. From a system's perspective, the focus lies heavily on sensors and how they can detect features from the hands. This latter perspective typically does not accommodate the user with intuitive gestures.

A system's perspective

In many gesture interfaces, the gestures are designed to replace GUI commands. The resulting gesture sets are highly idiosyncratic and are often hard to learn by novice users. Quek [172] designed 15 gestures that describe space and that specify spatial quantities. For space, these gestures capture tasks such as point, continue, stop, rotate, roll, pitch and track. Other gestures specified spatial quantities: large, small, left, right, up, down, farther, nearer. In these works no common interface tasks such as select, open and activate are formulated. Figure 3.5 depicts the gesture vocabulary from Quek [172] that was designed to describe space and specify spatial quantities for 3D drawing.

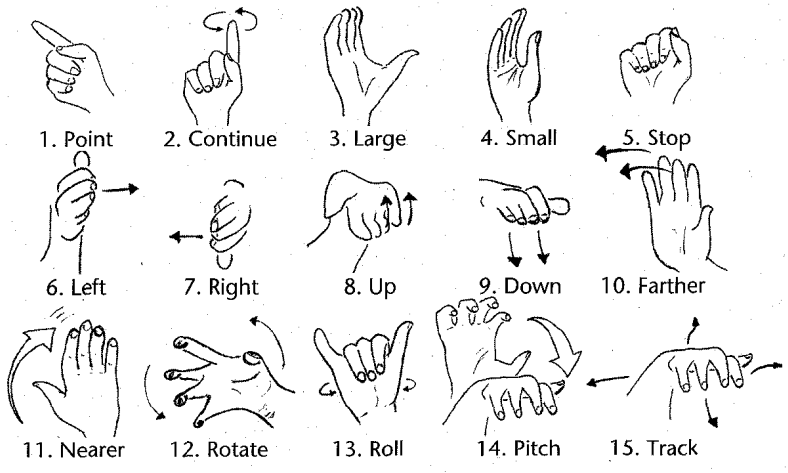


Figure 3.5: Gesture vocabulary by Quek designed to describe space and specify spatial quantities. Image by Quek [172], copyright 1996 IEEE.

Kavakli and Jayarathna [103] identify 32 gestures due to the limitation of their sensor set-up: "Up to 32 gestures can be defined, since there are 2^5(=32) possible combinations using flexure values fully flexed (<10%) and closed (>90%) for each finger. But all those gestures cannot be imitated by a human hand due to its physical restrictions." Kavakli *et al.* [104] further explored these gestures in a

working prototype. Their free-hand sketching application, DesIRe, was used to construct 3D drawings by directly observing and reacting to both hands. The DesIRe system included 29 gestures that transition between states and manipulate the 3D mesh. Gesture recognition was hard coded: bending a finger past a threshold value changed the appropriate phalanx-sensor from 0 to 1. Both the lack of any visual representation of the gesture set in combination with seemingly random gesture-task combinations makes this idiosyncratic gesture set hard to learn for end-users. In a similar pattern recognition approach, Lee and Kim [122] chose ten frequently used browsing commands of PowerPoint and assigned a gesture to each of them. Tasks such as moving between slides, starting and stopping the presentation were included. The gestures consisted of movement trajectories of a cursor. A similarly hard to learn gesture interface is described by Schlattman and Klein [187] who defined several hand shapes to represent tasks such as pointing or clutching. These hand shapes could be detected by an elaborate multi-camera system by virtue of one or two fingers protruding in the left, top or bottom sides of the detected hand shape. Tse *et al.* [205] built a gesture interface on top of the existing strategy game Warcraft 3. Using a DiamondTouch tabletop [35], Tse *et al.* required users to mark bounding boxes with two hands and issue commands to the selection by speech. Other gestures could not be implemented because of the lack of support from the DiamondTouch for disambiguation of two or more touches.

The SixthSense prototype is a mobile gesture interface that uses a projector instead of a display to visualize information on any surface [135]. The gestures in SixthSense, see Figure 3.8, are based on popular multi-touch systems and the Apple iPhone. The gesture set focuses on WIMP-like interfaces through pointing by ray-casting and selecting through button-up and button-down hand shapes with the thumb protruding from or enclosed in the hand respectively.

A human's perspective

Cutler *et al.* [31] built a responsive workbench that allows natural manipulation of virtual 3D models with both hands. Their tabletop system rear-projected the 3D models while two PINCH datagloves (see Section 2.4.4) were used to detect one-handed and two-handed gesturing. Three types of gesture-task combinations were defined: unimanual, bimanual symmetric and bimanual asymmetric. Guiard [65] was the main inspiration for manipulating objects bimanually by dividing tasks between the two hands. Cutler *et al.* [31] found that users combined two otherwise independent one-handed tools in a synergistic fashion. State-transitions between interface tasks mainly occurred by picking up physical tools, for example, a magnifying glass, to switch the system's state to 'zooming'. These physical tools made the interface states explicit for the user. Focusing on hand movements, Grossman *et al.* [63] created 3D curves with two hands by pressing buttons on hand-held spatial position sensors. By combining the movements from two hands, detailed curves could be produced that were impossible to create with one hand, see Figure 3.6.

Nielsen *et al.* [146] introduced design criteria for the human-based approach in which gestures need to be easy to perform and remember, intuitive, not physically

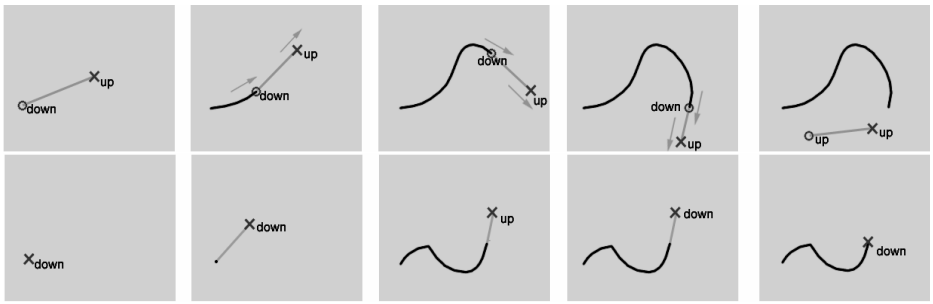


Figure 3.6: Digital tape drawing for describing 3D curves with two (top row) and one (bottom row) hands. Image by Grossman *et al.* [63], © 2002 ACM, Inc.

stressing and logical in terms of functionality. For moving objects in a paper mock-up with three different scenarios, Nielsen *et al.* found iconic gestures such as drawing a square to represent objects (e.g., a card) and, for selecting, they found pointing with an index finger to the object or by waving the hand in the general direction of that object. Other tasks such as move or select all required an explicit state-transition gesture that resulted in rather obscure gestures such as stopping an action with a 'halt' emblem, much as in [214]. These signal gestures are explicit and potentially intuitive for the users. However, they are complex which makes it hard to learn so that the gap in the gulf of execution might actually widen rather than close.

Beringer [8] tried to discover a gesture set for controlling his SmartKom system in a Wizard of Oz setting. Users pointed with one or more fingers and with one or two hands. Selecting was done by circling around an object or region while new forms of interactions such as 'no' or 'go back' were realized by a kind of waving of the hands.

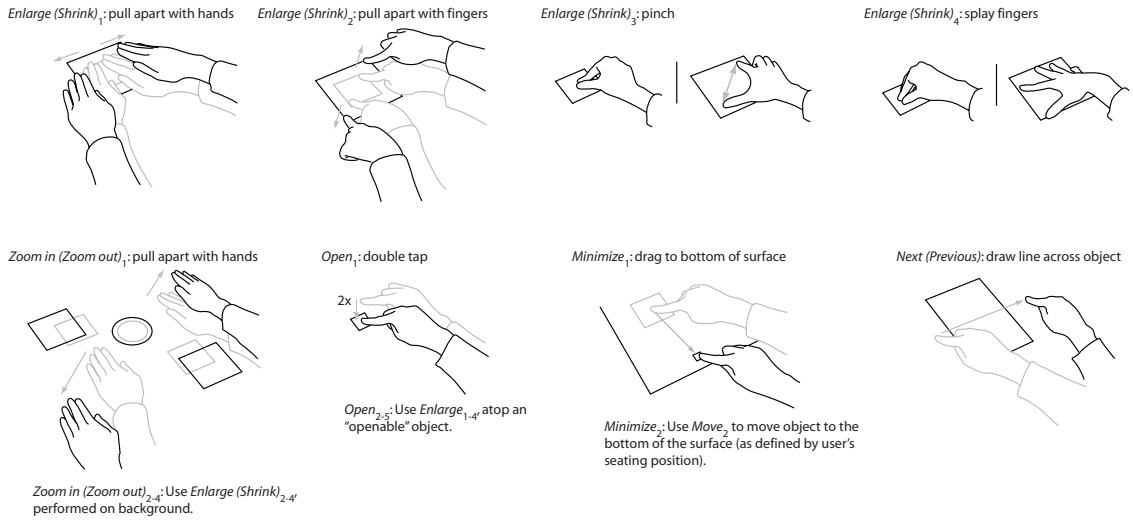


Figure 3.7: Some gestures from the user defined gesture set for multi-touch tabletops. Image by Wobbrock *et al.* [230], © 2009 ACM, Inc.

Hauptmann [77] found a surprising uniformity in the way that users communicated through speech and gesturing. He concludes that there are indeed intuitive,

common principles in gesture communication. Also, Hauptmann argues there are no expert users of gesture communication and that this channel is equally accessible to all computer users. In manipulating a cube, Hauptmann observed that complex bimanual movements were used but he did not list the actual gestures that his users made. Wobbrock *et al.* [230] applied a teach-back experiment to discover which gestures were made in multi-touch tabletop interaction when the result of the gesture was shown to the user. Figure 3.7 depicts a part of the resulting gesture set for multi-touch interaction. In their attempt to elicit gesture input for tabletop interaction from users, Wobbrock *et al.* [230] found that two hands are used for enlarging and zooming into an object but not for shrinking or minimizing. Wobbrock *et al.* found that users preferred to use only one hand so that their gesture set contains 31 unimanual gestures and 17 bimanual gestures, see Figure 3.7. There was some overlap between unimanual and bimanual gesturing, for example, with resizing where two fingers from one hand moved apart or where two hands move apart.

Sowa and Wachsmuth [195] designed a prototype system that can build a static spatial description of a virtual object based on the dynamic movements of two hands gesturing the spatial shape of the object. The speech and gesture utterances were decomposed into spatial entities and their relationships. In this sense, there was no construction of a gesture set for controlling an interface but rather understanding of the relationships between spatial gestured movements that, added up, represent and identify an object in the interface. Similar studies aimed to design organic 3D shapes [186] and common household objects such as a water bottle [87].

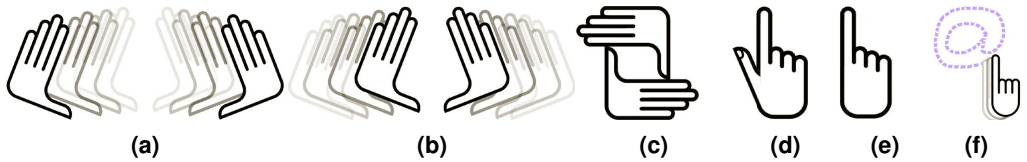


Figure 3.8: Gesture set of the Sixth Sense mobile gesture interface. (a) zoom in, (b) zoom out, (c) frame, (d) button-up, (e) button-down and (f) in-the-air drawing. Images by Mistry *et al.* [135], with permission.

3.4.2 Commercial gesture-based products

In contrast with experimental gesture interfaces, commercial interfaces focus far more on the end-user. If the interface does not improve existing interfaces, for example WIMP or iPhone, it will not sell. We distinguish between gesture interfaces meant for entertainment, access to complex visualizations and (fictional) interfaces seen in (science fiction) movies.

Entertainment interfaces

Mid 2007, Apple launched[2] the first version of its iPhone. The iPhone was the first mobile phone to introduce multi-touch technology to the public. Although being limited at first in its software to detecting one or two touches, the iPhone introduced intuitive gestures that allowed users to interact with detailed data through easy and predictable navigation (moving one finger), zooming (pinching two fingers together or moving them apart), and so on. The patent for mobile multi-touch filed by Apple in 2006 limits but does not prevent competitors to build multi-touch mobile phones of their own [85].

The Canesta[3] interface controls TV sets through gesturing. Three gestures control activating the TV (waving), switching channels (move hand left/right repeatedly) and changing the volume (moving the hand up/down repeatedly). A similar interface has been conceptualized by Bang and Olufsen in which users change the volume by tilting the physical remote and switch channels by waving their finger up/down in a hole inside the remote [210].

The gaming industry is also catching on to using gesture interfaces. The Nintendo Wii made gesture interfaces, through a hand-held controller (Wiimote), available to the general public[4]. The Wiimote is equipped with accelerometers that allow developers to detect and recognize motion trajectories of the hands. A broad range of games now makes use of the capabilities of these sensors: from throwing a fishing line to sword fighting to reloading shotguns, the in-game movements correspond to the actual act. Other game console manufacturers have introduced similar techniques. Microsoft's Xbox360 recently announced 'Project Natal'[5] that aims to allow computer vision-based detection of body pose: "If you know how to move your hands, shake your hips or speak you and your friends can jump into the fun – the only experience needed is life experience." The gestures mimic those in real life so kicking a ball requires the user to kick in midair. Toshiba's SpursEngine[6] includes a rudimentary gesture recognizer that analyses video images from an on-board laptop webcam. Three hand shapes are recognized: fist for pointing, fist with thumb-up for selecting and open hand for stopping.

Interacting with data visualizations

The g-speak spatial operating system by Oblong Industries[7] uses passive marker tracking on worn gloves to detect the hand location and shape of both hands. By combining hand poses and hand movements in synchronous or asynchronous and symmetrical or asymmetrical ways, the system performs requested tasks. The g-speak system has been used to implement g-stalt, an interface to manipulate com-

[2] `http://www.macworld.com/article/54769/2007/01/iphone.html`, June 16th, 2009.

[3] `http://canesta.com/`, October 13th, 2009.

[4] USA today, "Wii finds home in retirement communities, medical centers", May 14th, 2008.

[5] `http://www.gametrailers.com/video/e3-09-project-natal/50014`, June 4th, 2009.

[6] `http://www.tacp.toshiba.com/news/newsarticle.asp?newsid=191`, June 16th, 2009.

[7] `http://oblong.com`, June 16th, 2009. MIT spin-off company based on their work with the 'Minorty Report' video prototype.

plex data sets with the hands. The gestures in g-stalt are tuned to manipulate photos, see Figure 3.9. Although the used g-speak gesture set is not described in detail by Oblong, video clips of the functional system show users pointing through ray-casting, clicking through a *ThumbTrigger* gesture [64] and grabbing the whole screen by pinching the thumb and index finger of both hands. Although the g-speak system and its g-stalt implementation seem imposing, the gestures require users to hold their hands in midair during the complete interaction. Although this follows to the KC-model [65], by letting the NP-hand set a reference frame for the P-hand, it is likely to result in fatigue or 'gorilla arms'[8] rapidly because the hands have no position to rest.

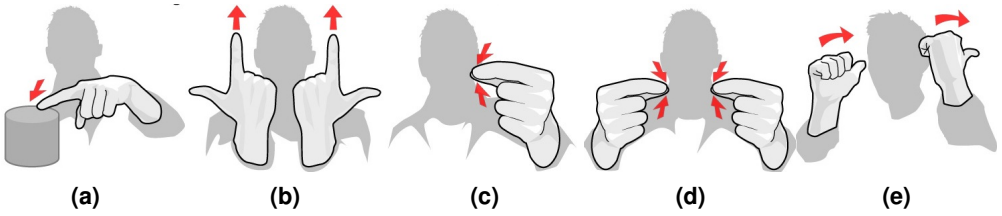


Figure 3.9: Gesture set from the g-stalt photo manipulation interface. (a) 'get photos of this object', (b) 'bring up more photos', (c) 'grab and move space', (d) 'grab and rotate space' and (e) 'reset view'. Images taken from Zigelbaum's website[9].

In our own research [44], interactive tables have also proven to be extremely suited for collaboration while users stand or sit around the table, sharing a common display [51]. We have evaluated collaborative behavior with a biological visualization tool that fills the table with a 3D molecule. Bioinformaticians stand around the table to discover function from form. We explored various gestures to control the resulting multi-user, multi-touch interface. We ended up with the following gestures: a fist performed a reset-action, dragging with one or two fingers rotated the molecule, dragging with the whole hand translated the molecule, dragging two fingers some 10 cm apart towards the user opened a context menu and moving two fingers apart or together zoomed in and out of the molecule.

Fictional interfaces

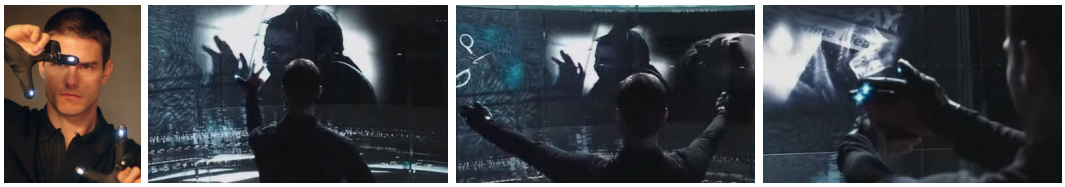
In 1994, Tognazzini [203] described Starfire, a video prototype project that showed a day in the life of a typical Starfire user in the year 2004. Both large and small displays were portrayed in fictional interfaces. The hands manipulate and move objects around in a physically believable way [1]; touch was not required per se. In the making of Starfire, it was surprising to observe that the actress gestured too quickly. The resulting response in the video-prototyped interface was unintended

[8]Humans are not built to hold their arms at waist or head height for extended periods of time. After more than a very few actions, the arm begins to feel sore, cramped, and oversized. The user looks like a gorilla while using the touch screen and feels like one afterwards. Gorrila-arm is shorthand for 'how is it going to fly in *real* use?'.

[9]`http://zig.media.mit.edu/Work/G-stalt`, June 16th, 2009.

throwing of the virtual object across the display instead of moving it, attached to the actress' hand[9].

The g-speak developers were involved in the creation of the science-fiction film Minority Report[10], see Figure 3.10. In Minority Report, actor Tom Cruise controls large vertical display with actively marked gloves by pointing through ray-casting, zooming by moving the hands relative to/from each other in depth and selecting by encircling a target. In Paycheck[11], actor Ben Affleck manipulates a 3D hologram representation of a product design. Through two hand-held pens, Affleck can mould the virtual design, using button presses for selecting.



(a) (b) (c) (d)

Figure 3.10: Gestures as seen in the Science Fiction film Minority Movie. Tom Cruise controls a vertical screen with (a) actively marked gloves for (b) selection, (c) positioning and (d) zooming.

Tabletop interfaces are increasingly popular in science fiction films as well. However, these interfaces lean heavily on real world interactions. After all, a display surface that can sense touch, selecting items by tapping with your finger or a pen is immediately appealing, as it mimics real world interaction. Tangible and virtual objects alike are manipulated in similar ways by moving and rotating them on the surface. Examples of touch sensitive displays are seen in the James Bond movie Quantum of Solace [12] in which the MI5 team is analyzing fake money bills and in the movie The Island[13] in which actor Ewan McGregor draws an artist impression of a boat which is then pondered upon by himself and his psychotherapist.

3.4.3 Research agenda

Intuitive gestures in most experimental and commercial gesture interfaces are designed to be easy to learn, easy to remember and, in most cases, easy to perform. However, this does not mean that these intuitive gestures come naturally. The gestures that we have described in this chapter show a lot of overlap between products, movies and research projects. We list the most frequently occurring gestures for the four elementary interface tasks that we have defined (see Figure 2.3): #0; out-of-range, #1; tracking, #2; selected and #3; manipulating, for example, resizing, activating and positioning. For tracking, two alternatives

[9] `http://asktog.com/starfire/starfirescript.html`, October 1st, 2009.
[10] `http://www.imdb.com/title/tt0181689`, June 16th, 2009.
[11] `http://www.imdb.com/title/tt0338337.`, June 16th, 2009
[12] `http://www.imdb.com/title/tt0830515`, June 16th, 2009.
[13] `http://www.imdb.com/title/tt0399201`, June 16th, 2009.

were found, ray-casting ([8; 187; 172; 214], Minority Report and SpursEngine) and tapping repeatedly with the whole hand ([172]). For selecting, there were very diverse solutions: dwell (Minority Report), tapping with either index or thumb ([40; 64; 135; 186; 214; 215; 230] and g-stalt), pinch ([31; 63]) and marking a bounding box or circling objects ([8; 135; 146; 205]). There are numerous implementations for manipulating. We describe the most frequently occurring tasks that are performed through gesturing. For zooming in and out, two fingers or hands are pinched or moved apart ([135; 230], iPhone, Paycheck, The Island) or two hands are moved apart/together in depth (Minority Report). Stopping or closing an action is done with a fist ([172]), a flat hand as in 'halt' ([214; 215], SpursEngine), move an object to a deactivation area ([214; 230]) and moving the hands over a shoulder with the thumbs up (g-stalt). Switching to next or previous was done by moving a finger or hand in a direction, often left for previous and right for next ([230], Canesta TV, Bang & Olufsen). Also, to describe spatial shapes, the fingers/hands indicate (relative) distance and shape by moving apart ([135; 172; 195]) or by following (parts of) the spatial shape of the intended object ([63; 87; 186]).

3.5 Summary

This chapter described two gesture taxonomies. First, we provided an overview of how gestures are categorized as seen from more traditional gesture research in which the relationship between gesture and speech is the focal point. Six gesture categories are used in that area of research: adaptor, emblem, deictic, iconic, metaphoric and beat gestures. From another point of view, that of human-computer interaction (HCI), we categorized gestures as deictic, manipulative, gesticulation, semaphore and sign language gestures. There is significant overlap between the definitions in both taxonomies. The main difference follows from the application area: the traditional taxonomy classifies gestures that accompany speech while the HCI taxonomy classifies gestures that can be used for interacting, in various ways, with computer systems. In the remainder of this work we will look at adaptor and emblem gestures, as seen from the traditional taxonomy, and at manipulative and semaphoric gestures, as seen from the HCI taxonomy. The common denominator is that the gestures manipulate objects and that they do not require information that is communicated through speech.

Gestures can be divided into the preparation, stroke and retraction phases. Recognizing gestures typically focuses on the stroke phase in which contains the most explicit finger, hand and arm movements. The stroke normally provides the most information in a gesture. We have shown how gestures have been implemented in human-computer interaction dialogues; typically as a gesture set that the user has to learn in order to operate the system. This gesture set is often based on the sensors that are used to look at the user gesturing and not, as is the focus of this thesis, on finding the appropriate sensor for the gestures that come naturally to the user.

Intuitive gestures come naturally to the user, arguably because of manipulating objects in every day life or from gesticulating in human-human communication. In

addition, other objects from every day life are by now familiar human-computer interfaces, for example, WIMP systems. These interfaces have indoctrinated the user to accept gestures to be intuitive and natural. It will be interesting to discover which gestures come naturally to uninstructed users of a gesture interface. Such a study will have to take into account the reasons why the users made the gestures they did and why they thought those gestures are natural. Also, it is interesting to find what users think in terms of intuitiveness and physical effort involved of the gesture solutions that are presented in existing systems, especially given the large overlap in those gesture solutions.

Part II

Experiments

Chapter 4

Uninstructed Gesturing

> "[..] media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers groceries causes change in our nutrition."
>
> **Richard Clark**
> [29, p.2] *Learning from media - arguments, analysis, and evidence*, vol. 1 of *Perspectives in Instructional Technology and Distance Learning*, pp. 1Ű12. Information Age Publishing: 2001.

This part describes a series of experiments with which we explore gesture-based interaction with large displays that cannot be touched. We first explore, from a perspective of human behavior, which gestures are suited to control large displays. For that, we performed an experiment in which we asked subjects to gesture as they saw fit. Their goal was to issue commands to the interface with only their hands and without being told how to gesture. This experiment is described in this chapter. Second, we try to understand why these gestures are suited to control large displays. We performed a large-scale experiment in which multiple gestures from our first experiment, fictional and fictitious sources were evaluated for a series of elementary interface tasks. This experiment is described in Chapter 5. Third, we validated our findings from this large-scale experiment with a partially working prototype interface, see Chapter 6. This part is concluded by describing a qualitative evaluation of a fully operational gesture-based interface. With this fourth experiment, described in Chapter 7, we explore how a gesture-based interface such as the one portrayed in Minority Report can be made a reality.

4.1 Introduction

To discover which gestures are considered natural by potential users, we asked uninstructed users to produce gestures that they felt would match a given command. This Wizard of Oz experiment was meant as an exploratory study with the goal to gather a list of gestures that come naturally. In a Wizard of Oz approach, the user is led to believe that she is in actual control while, in fact, an operator is controlling the interface. In a similar study, Hauptmann [77] asked users which gestures

they would make to spatially rotate, translate and scale three-dimensional graphic objects on a display with just their hands. Users moved their hands freely in this ethnographic study. Hauptmann focused mainly on the number of hands and fingers that were used to interact but not on the actual hand shapes that were used. To discover which gestures users would make to control multi-touch applications, Wobbrock *et al.* [230] asked their users to teach-back to the developers how the system works. Their users where given a limited number of tasks for which they needed to explain how it worked to the developers. The teach-back approach used by Wobbrock *et al.* is quite similar to the Wizard of Oz approach used by Hauptmann [77]. Beringer [8] applied a Wizard of Oz set-up where users were asked to interact with a digital shopping window through gesturing. In order to prevent their users from regarding the application as just another Windows-style application, an entirely different look-and-feel was introduced. Beringer found that their users felt somewhat inhibited in interacting with their Wizard of Oz system due to the lack of introductoral information about the possible gestures. However, despite this inhibition, it was found that many users did interact though gesturing and that they even used new forms of gestures.

This chapter is structured as follows. In Section 4.2, we describe the method that was used in this Wizard of Oz experiment. Uninstructed participants were asked to manipulate a topographical map by gesturing with their hands. Section 4.3 reports our findings that are based on video-analysis of the trials using Anvil [110] and ASCII Stokoe[1] for annotating the observed gestures. Section 4.4 then sums up the gestures that were made by uninstructed users of a gesture interface. Concluding this chapter, we discuss our findings in Section 4.5.

4.2 Method

The goal for this exploratory study was to discover which gestures are made for issuing a set of simple commands without instructing the participants *how* to gesture. We chose a Wizard of Oz setting in which an operator was in actual control of the interface while allowing our participants to believe that *they* were in control. In this set-up, there was no need to build a functional interface that interpreted participants' gesturing: the operator performed that task. Our participants were asked to manipulate a topographic map of our university's local surroundings (Twente, the Netherlands). These participants were not expected to have knowledge of the local topography as it was unimportant to complete the assignments as fast as possible. Participants could issue two commands to complete each assignment successfully: panning and zooming. These commands both represent different implementations of the manipulation state in our four-state model for input, see Section 2.3. The implementation of our four-state model for this study does not include the tracking state (#1) nor the selecting (#2) state. The user is either not interacting when out-of-range(#0), panning (#3) or zooming (#3). Note that the user moved directly

[1] `http://www.speakeasy.org/~mamandel/ASCII-Stokoe.txt`, June 4th, 2009.

from out-of-range to one of the two implementations of manipulating; panning and zooming.

Figure 4.1c shows the map and the Wizard of Oz set-up that was used. Participants stood on a marked square at three meters from a 400 × 400 cm projection screen on which the map was projected with a total resolution of 1600 × 1200 pixels. To avoid users thinking that the system is just another Windows-style application, we followed Beringer [8] to give the application a different look and feel: the map was projected full-screen with no further interface components. In this manner we hoped to prevent users from reverting to interactions that they were already familiar with, such as using simple 'click' gestures as they would do with the mouse. Three increasingly difficult assignments were given to the participants in which the map view needed to be moved to a specific location on a specific zoom level. Assignments consisted of locating and displaying one or more specific town(s) and of positioning the view port so that these target(s) would fill the screen entirely. Participants were thus required to move the map and change the zoom level to complete each assignment successfully. Whenever the participants were searching for a location on the map for more than 2 minutes, we hinted the direction in which the goal was relative to the current view port location.

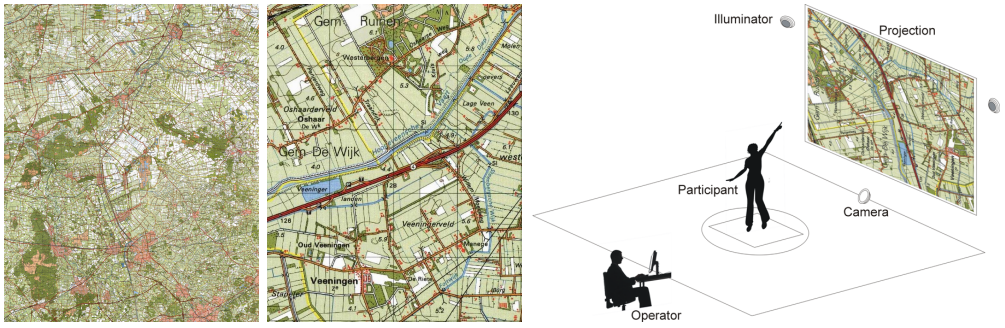


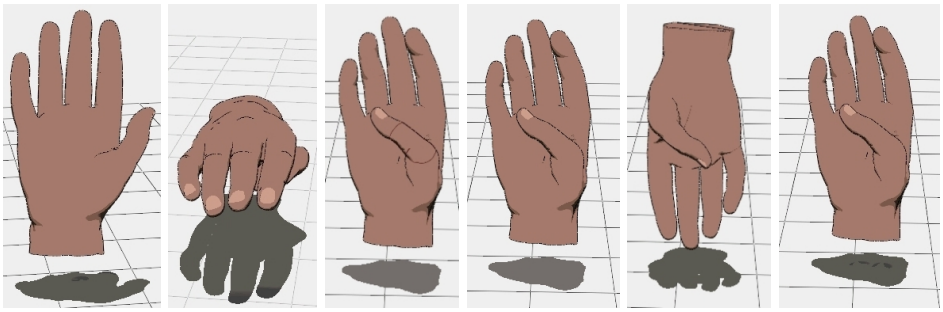
Figure 4.1: The map that the participants had to manipulate through gesturing: (a) the whole map, (b) top-left map detail and (c) the set-up of this experiment.

Each session began with a brief explanation for the participant. The wizard, or operator, was introduced as a technician who would perform minor adjustments to a working gesture interface during a brief speak-out-loud phase at the beginning of each trial. Note that during this phase, the operator chose the couplings of gesture and state-change, including when a participant was in the out-of-range state. The operator was instructed to respond only to hand movements and to ignore any verbal commands except during the brief speak-out-loud phase. The spoken dialogue in the speak-out-loud phase tuned gestures and their intended system response. The system response was limited to visual feedback with the map either panning or resizing as instructed by the participant's gesturing. After finishing the three assignments we briefly interviewed the participants to question them about their motives, if there were any, for choosing specific gestures for performing pan or zoom in/out.

4.2.1 Video annotations

Each trial was videotaped with a camera facing the subject, see Figure 4.1c. As the room had to be darkened so that the projection would be sufficiently bright we illuminated the scene with a pair of infrared lamps. The light was invisible for our subjects but it did improve the scene illumination in our recordings noticeably. We omitted sections of the video recordings that were deemed unintended communications, for example, turning around to face the wizard asking him to adjust the system's responses. The recorded trials were annotated in ASCII Stokoe and using Anvil [110]. ASCII Stokoe, like HamNoSys [170], is designed for sign language annotations but its annotations are, unlike HamNoSys, not so complex, because ASCII Stokoe uses only ASCII characters. We expanded the annotation with additional symbols $\&$ and Z that represent hand movements with the same hand shape and orientation respectively. In addition, we added the circumfix symbol $\mathrm{S}(..)$ to represent synchronized bimanual movements. Note that asynchronous movements can already by adequately described in Stokoe. A brief pilot session led us to believe that for our limited set of tasks no asynchronous bimanual gestures could be expected.

The ASCII Stokoe annotations were at first so detailed that there was little overlap between the observed gestures even though the operator had interpreted some gestures to have the same meaning. We abstracted our annotations based on the assumption that similar gestures would have a similar meaning [8]. For example, differences between hand orientation (slightly upwards compared to fully upwards) and hand shapes (cupped hand versus slightly stretched hand) are grouped together. In ASCII Stokoe notation we grouped $\mathrm{B5}$ with Bb, fB, B, $\mathrm{B}\wedge$ and Bv, see Figure 4.2. Also, we grouped $\mathrm{Q/C/f}$ (away from the signer) and $\mathrm{Q/C/t}$ (towards the signer), in addition to $\mathrm{Q/C</\wedge}$ (upwards) and $\mathrm{Q/C</v}$ (downwards) for directions in panning. We grouped A (fist) and G (pointing hand) in the zoom task when users moved their two hands apart, for example, $\mathrm{S(Q/C/\#\{A\}\ Q/A/Z\ Q/A/]\{C\})}$ where the distance between the hands was important, not the hand shapes, see Figure 4.6a.



(a) $\mathrm{B5}$ **(b)** Bb **(c)** fB **(d)** B **(e)** Bv **(f)** $\mathrm{B}\wedge$

Figure 4.2: ASCII Stokoe hand shape abstractions.

4.3 Results

Nine participants took part in our within-subjects design. They were on average 27 years old ($\sigma = 6$), ranging from 19–36 years. Seven participants held a BSc's degree, two had completed a Master's degree. One participant was female, eight were male. All participants were right handed, no participant was ambidextrous. Figure 4.3 depicts our participants' proficiency with human-computer interactions that may have influenced their choice for gestures. Participants in our Wizard of Oz experiment were proficient with computers ($\bar{x} = 2.8$, $\sigma = .4$), the internet ($\bar{x} = 3.0$, $\sigma = 0$) and map applications ($\bar{x} = 2.8$, $\sigma = .4$). They were somewhat familiar with the topography of the Twente region in the Netherlands ($\bar{x} = 2.1$, $\sigma = .6$) but not very proficient with computer games ($\bar{x} = 1.4$, $\sigma = .5$) and 3D drawing applications such as CAD/CAM ($\bar{x} = 1.7$, $\sigma = .5$).

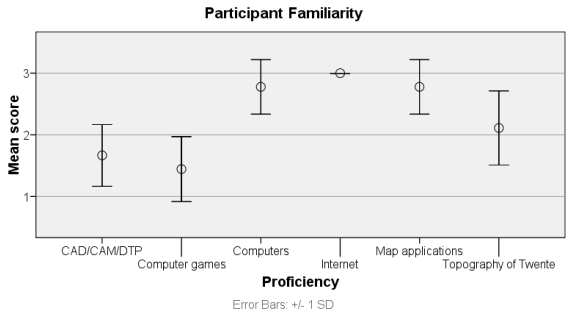


Figure 4.3: Participants' proficiency with similar interactions before the experiment on a 1 (low) to 3 (strong) scale.

Subject	Assign. 1	Assign. 2	Assign. 3	# pans	# zooms
1	0'48"	0'36"	0'47"	42	17
2	1'15"	0'54"	0'40"	96	4
3	2'31"	1'20"	0'42"	51	74
4	1'35"	1'25"	1'29"	62	33
5	0'39"	0'22"	0'27"	37	12
6	0'56"	0'55"	0'17"	27	20
7	0'53"	0'50"	0'48"	33	7
8	1'00"	0'44"	1'07"	35	9
9	1'41"	0'52"	0'57"	26	17

Table 4.1: Assignment completion times (m'ss") and the total number of gestures that were made to finish the assignments.

The completion time for the assignments was on average 1'15" minutes for assignment 1 (locate one town, $\sigma = 35s$), 0'53" minutes for assignment 2 (locate two towns, $\sigma = 8s$) and 0'48" minutes for assignment 3 (locate three towns, $\sigma = 7s$). We annotated this video data and made gesture abstractions from it. Table 4.1 shows

the average task completion time and the number of pan and zoom gestures that were made per participant. After some initial abstraction we identified 14 distinct pan and 13 distinct zoom gestures. By generalizing over annotations, as described in Section 4.2, we could further reduce the number of gesture distinctions to leave 3 distinct pan and 6 distinct zoom gestures. Typically, participants panned twice as often as they zoomed, see Table 4.1. Figure 4.4 displays the number of occurrences per gesture per assignment.

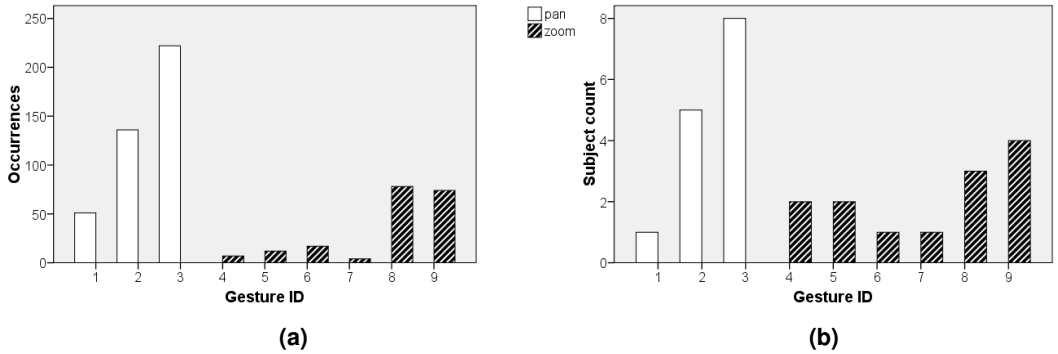


Figure 4.4: Gesture occurrences per assignment. (a) The total count of a gesture and (b) the number of subjects that gestured identically.

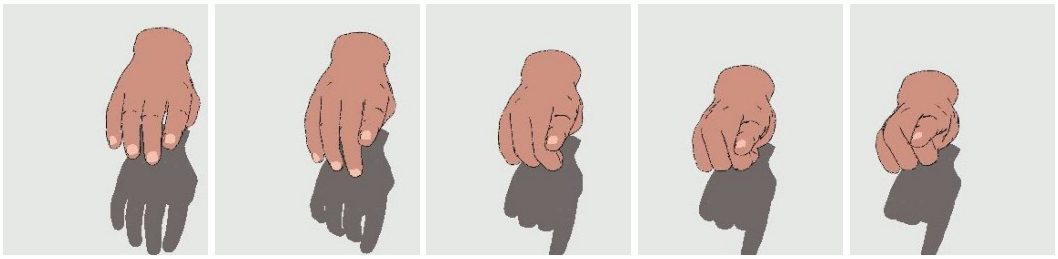
Uninstructed panning

For the pan task, two gestures (IDs 2 and 3 respectively, see Figure 4.5) were observed to occur significantly ($p = .01$) more often than other observed gestures. In addition, these gestures were observed in most users. The difference between these two gestures is that in gesture 2 the hand will be closed at the beginning of the movement and opened at its end as if to grab and release the canvas. In gesture 3 these changes in hand shape do not occur. The other gesture is largely similar to gesture 3 except that the hand is closed with the index finger extended (pointing).

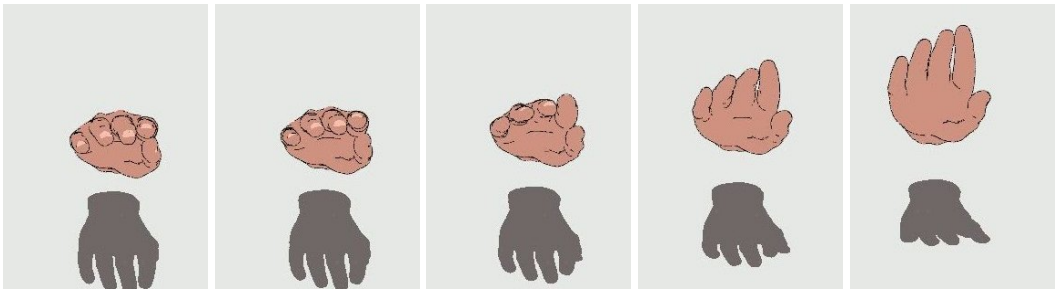
Uninstructed zooming

For the zoom task, we observed a similar distinction; gestures 8 and 9 (see Figure 4.6) occur more ($p = .07$) for zooming than the other four observed gestures. These two gestures are, in fact, very similar still; differing in hand shape only as in the pan task. Gesture 8, like gesture 2 for panning, grabs and releases the canvas while gesture 9 will explicitly stretch the hand during the zoom movements. Figures 4.5 and 4.6 illustrate the differences between the most occurring gestures for panning and zooming. A distinction can be made in the type of zoom gestures based on using one or two hands: gestures 4, 5 and 6 use one hand and 7, 8 and 9 use both hands. The number of subjects who chose to use two hands (5 subjects) does not differ that much from the subjects using a single hand (4 subjects). Subjects were

consistent in their choice for using either one or two hands; no participants used both gestures.



(a) gesture 2: Q/Mf/d (136 occurrences in 5 subjects)

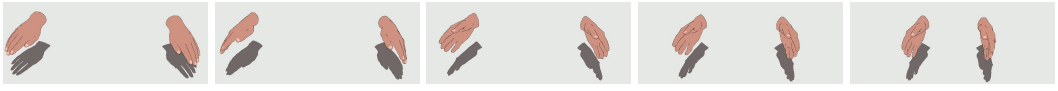


(b) gesture 3: Q/C/],f{B} Q/B/& Q/B/#,t{C} (222 occurrences in 8 subjects)

Figure 4.5: The two most occurring gestures for panning (ASCII Stokoe notation in gray is not depicted): (a) pointing hand away from the user towards the display that moves around, relaxing the hand would release and (b) relaxed/cupped hand to stretched hand that moves around for panning, relaxing and retracting the hand would release.

(a) gesture 8: S(Q/C/#{A} Q/A/Z Q/A/]{C}) (78 occurrences in 3 subjects)

(b) gesture 9: S(Q/C/],f{B} Q/B/Z Q/B/#,t{C}) (74 occurrences in 4 subjects)



(c) gesture 9: S(Q/C/],f{B} Q/B/Z Q/B/#,t{C}) (74 occurrences in 4 subjects)

Figure 4.6: The two most occurring gestures for zooming (ASCII Stokoe notation in gray is not depicted): (a) relaxed/cupped hands to pointing hands that move apart for zooming out and move together for zooming in, relaxing and retracting the hands would release and (b) relaxed/cupped hands to stretched hands that move apart for zooming out and move together for zooming in, (c) relaxing and retracting the hands would release.

4.4 Conclusions

With a Wizard of Oz experiment we explored uninstructed gesturing by users who were asked to control a map through panning and zooming in and out. Users were videotaped and these recordings of our subjects gesturing were annotated in ASCII Stokoe using Anvil [110]. Although this annotation approach can adequately describe all observed gestures, unless it can be automated we argue to use some other method of writing down gestural movements.

A total of 14 pan and 13 zoom gestures were counted based on the initial annotations. In our annotations, we grouped hand shapes and movements that had similar meaning together. This led to three distinct pan and six distinct zoom gestures that differ in annotation but not much in meaning. By grouping together gestures with similar annotations we abstracted the different gestures somewhat to reduce the number of different gestures for pan (3) and zoom (6). By counting the occurrences of these gestures we identified two gestures for both the pan and zoom tasks. We found great uniformity between the gestures made by our subjects. In addition, the main difference between the gestures for the pan and zoom tasks seems to be the hand shape used in the gesture preparation and retraction phases, not the movement made in the gesture stroke.

The gestures that we observed differ mostly in the preparation (start) and retraction (end) phases of the gesture phrase [134]. For both pan and zoom tasks, our subjects explicitly marked the start and end of their gesture by changing their hand shape from rest to a flat hand or pointing hand for panning and two flat or pointing hands for zooming. In addition, we found that in the stroke phase the movements can more or less be directly used as parameter changes for panning and zooming; the hand shape during these movements does not matter much. Previously it has been found that the number of fingers used to perform a gesture does not matter for the gesture's meaning [230].

Our subjects consistently apply the same idiosyncratic combinations of gesture and command with a great deal of similarity between users. This leads us to believe that it is possible to construct a more complete set of gesture-commands for large display control that comes naturally to the users.

4.5 Discussion

The set-up used in this experiment raises some questions with respect to biases on our results. First, the operator chose how to respond to the users, by also listening to speak-out-loud explanations. However, it is unknown how other operators would respond in the same setting. Also, the consistency with which an operator responded to users gesturing may not have always be adequate. For example, during one trial in the pilot session, our operator mistook zooming out for zooming in. It is insightful that the user quickly adjusted himself to fit the actual response of the system instead of retrying. We believe that this response is caused by the prototyped set-up of this experiment: users readily accepted minor errors in the

interaction. Because the task set was quite limited we do not expect that such interpretation mishaps have had a great impact on our results. However, observing users with more operators would have made the experiment more reliable yet much more time-consuming. Second, to what extent were our subjects uninstructed to begin with? Our subjects were proficient with computers, the internet and digital map applications. How much did these proficiencies influence the gestures that the subjects made? Some subjects remarked that their gestures were mimicked from the Apple iPhone that readily uses zooming in/out by moving two fingers apart/together. It is then no surprise to find that this gesture came up as one of the gestures for zooming. However, did it come naturally or was it instructed to some degree by the iPhone interface? Whatever the case, we believe that this question is no longer relevant. No matter its source, this influence is here to stay, as we also have repeatedly observed in our other work on multi-touch sensitive surfaces where users *expect* this gesture for resizing objects, even when they have never held an iPhone [44; 45]. We took effort to remove familiar look-and-feel components from the interface to prevent users from reverting to familiar interactions from, for example, the WIMP metaphor. However, the iPhone interface also does away with these familiar interface components. Thus, it might be possible that this is the cause that we mainly observed these two gestures—moving two fingers or hands apart—for the task of zooming in. However, we believe that it is rather the transparent system response rather than the lack of interface components that causes users to gesture in this way.

4.5.1 Retrospection

Subjects remarked, in interviews after the Wizard of Oz experiment was over, that their gesture choices were often based on their knowledge of 'mainstream' gesture interfaces such as the Apple iPhone or that they had mimicked movements that they remembered from science fiction movies such as Minority Report. We are unsure how this has influenced our results but it is clear that users readily accept such 'predefined' gesturing as a natural form of interacting.

Chapter 5

The Public on Gestures

> "Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution."
>
> **Albert Einstein**
> Theoretical physicist, 1879-1955 – What Life Means to Einstein, Saturday Evening Post, October 26, 1929.

Intuitive gesturing is gesturing that comes naturally. These gestures minimize the gap in the gulf of execution [152], so that the psychological language that describes the user's goals and the action-object language of an interface match best [12], see also Section 2.1. The previous chapter described a Wizard of Oz experiment in which we found out that users apply the same gesture for each command that they issue consistently and with great similarity between users. In this chapter we try to understand why these gestures are suited to control large displays. By consulting a large, uniform sample on what they find intuitive, we gain an insight into which gestures come naturally and which gestures do not. This sample is taken from our target audience of potential users of large displays. A large user group can think up more gestures than three experts can [230]. Clearly, such a large sample only describes which gestures that specific group finds intuitive. It does not describe whether other social, educational or cultural groups share this opinion. That is why we are also interested in the reasons why a gesture is considered intuitive or not. If a specific (part of an) interface, say the WIMP metaphor, is the reason for users to consider tapping an object to select it to be intuitive, it will most likely not be intuitive to those users that are unfamiliar with this interface. If a gesture is based on things that humans do in everyday life, for example, picking up an object and placing it on a table, it might appeal to other user groups as well. Again, the aim in this thesis is to evaluate gestures for explicit command-giving. In order to reach a large sample we chose to perform an online questionnaire in which we asked participants to rate gestures for a task based on intuitiveness, whether they would use it and the amount of physical effort that the gesture would entail. The aim of here is to select, based on consensus, a gesture set that is suited for explicit command-giving in large display interfaces.

This chapter is structured as follows. Section 5.1 introduces the design of our online questionnaire. In Section 5.2 we describe the scenarios that were included in our investigation. Scenarios consist of commands to issue in a gesture interface and the gestures that we have selected for these commands. We report our results per interface command in Section 5.3. Section 5.4 summarizes the results and Section 5.5 then describes the gesture set that we found to suit explicit command-giving to large displays. This chapter is concluded by a discussion on the results and implications of this online questionnaire and its design, see Section 5.6.

5.1 Online questionnaire design

In order to fully appreciate or dislike a gesture for issuing a specific command we argue that the user needs to experience it in a working system. Getting objective results then becomes cost-ineffective as opposed to gathering such information from a large sample online. However, in that case we must rely on the users correctly imagining how it would be to gesture as we show them. This will bias the results to an unknown extent. We expect that this bias is minimal but that will need to be proven by comparing the results to those from a working interface (see Chapter 6).

The online questionnaire is based on short videoclips that show the user issue commands to a mock-up interface through gestures. Each videoclip showed an actor issuing a command to a large display with a gesture. The display responded in a predictable way so that the participants can fully understand what the gesture would require them to do in such an interface [8; 203]. The videoclips for the assignments were shot in a Wizard of Oz setting with an operator controlling the display from behind the scene so that the resulting videoclip shows a seemingly operational gesture interface. The actor was filmed from over his right shoulder so that the display is clearly visible, as are both hands of the actor when they are not in the out-of-range state, see Figures 5.2–5.7. Each gesture was also briefly described textually, see Appendix A, and the videoclip could be run repeatedly. We asked participants to score (on a seven-point Likert-style scale) the intuitiveness (‘1: very difficult’ - ‘7: very intuitive’), the amount of physical effort that is required (‘1: little effort’ - ‘7: much effort’) and whether they would gesture in this way (‘1: no way’ - ‘7: for certain’). In addition, there was room for optional comments. Figure A.1 depicts the online questionnaire[1].

5.1.1 Abstract application

We made sure that the videoclips displayed both the actor issuing a command by gesturing and the display responding to the gesture. This approach was meant to help our participants, most of whom were inexperienced with such interactions, to imagine and appreciate the portrayed interaction [182]. To avoid users thinking that the system is just another Windows-style application, we followed Beringer [8]

[1] `http://fikkert.net/experiment.php`, October 8th, 2009.

to give the application a different look and feel. We hoped to prevent users from reverting to interactions that they are already familiar with, such as using simple 'click' gestures as they would do with the mouse. Windows- or Mac-style interface elements were avoided but even then our participants may have thought of the desktop paradigm when thinking about suitable gestures for a command [230]. An application that abstracts from any window-based interface was designed and built based on our four states, see Section 2.3. This abstract application displays one target that represents an icon or window object common in a WIMP interface. Note that this abstract application does not *do* anything, nor is it supposed to. The application merely serves as a means of visual feedback to the participants in this online questionnaire.

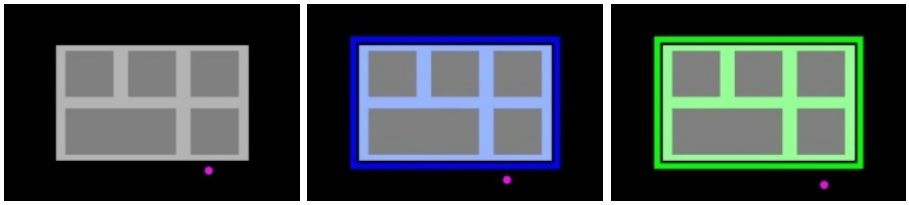


Figure 5.1: The three explicit states in our abstract application: (a) Pointing towards targets (pink cursor) (b) the target has been selected (blue edge and hue) and (c) the target has been activated (green edge and hue).

Figure 5.1 illustrates the three states that are explicitly visualized in our abstract application. Note that the out-of-range state is implicit: when the actor points to somewhere not on the screen, the application simply does not show the cursor. In the tracking-state, only the mouse cursor is showing (pink) on the screen. When the actor performs one of the select gestures, the application shifts to the selected-state. Moving back to the tracking-state is done by making a deselect gesture. Similarly, we move to and from the manipulation-state when a(n) (de)activate gesture or a zoom gesture is made. Note that there are two different implementations of our manipulation-state: zooming and (de)activating.

5.1.2 Analyzing the questionnaire

We first checked whether our data followed a normal distribution with a D'Agostino-Pearson K^2 analysis [33]. This is more robust and stringent than popular alternatives such as the one-sample Kolmogorov-Smirnov (K-S) and χ^2 analyses. Arguably, we might have used the Shapiro-Wilk analysis which works very well if every value is unique, but does not work so well when there are ties. Whatever analysis method is used, if we find a significant result we should examine our data for skewness, kurtosis or both because that could indicate a ceiling or floor effect.

Section 5.3.1 will first show that there was no normal distribution in our data. Therefore, the assumptions underlying a typical ANOVA analysis did not hold. We therefore had to apply the non-parametric alternative to ANOVA, Kruskal-Wallis H,

to assess whether there was a significant difference between gestures. The Kruskal-Wallis H analysis is more powerful than its alternative, the median analysis, because it takes rank size into account rather than only the above-below dichotomy of the median analysis. Kruskal-Wallis H only detects group differences so a post hoc analysis is required to identify these differences. We applied the Mann-Whitney U analysis (also know as the Wilcoxon-Mann-Whitney analysis) for further pair-wise comparisons. A Mann-Whitney U analysis is equivalent to a pair-wise Kruskal-Wallis H analysis.

5.2 Scenarios

The Wizard of Oz experiment that was described in the previous chapter was limited by the context of our map application. Only two commands could be issued: panning and zooming. In this investigation we include all four states that were described in Section 2.3: out-of-range, tracking, selected, manipulating. We chose to include zooming as an implementation of the manipulation-state because most literature on gesture-based interfaces focuses on manipulating images, often with a demonstrator application to resize, position and orient photos. Note that we did not include a gesture that changes only the orientation of an object, because that is often included in positioning or resizing an object with two hands [135]. We added activating and deactivating an object and a right-mouse-like command that would open/close a context menu. For each the commands that translate to a state-change in our model, we selected a number of frequently occurring gestures. See Section 3.4 for an overview of possible gesture sets. The commands were ordered in a predefined sequence because users would need to make up their mind first, for example, about how they would point before they could select. The commands that we presented to the user are, in this order: point, select, deselect, resize, activate & deactivate and open/close a context menu. The various gestures were completely randomized per command. On the following pages, these six commands and the various evaluated gestures per command are described in more detail. A series of snapshots from these videoclips is included to illustrate on what information our participants judged the clips.

Pointing

Pointing to a target is one of the fundamental functionalities in any interface as shown in [241]. This is the tracking-state (#1) in our four-state model. It is possible to point directly at a target, which translates to ray-casting, or to point indirectly which is often due to an input device that serves as a go-between, for example, a multi-touch table [73], mouse [128] or light pen [113]. Direct pointing, or ray-casting, has been described in Section 3.4 as the most popular gesture for pointing [8; 187; 172; 214]. Indirect pointing required tapping the hand repetitively in the direction where the cursor should move [172]. The gestures that were compared are:

1. *Ray-casting*: pointing to a location on the screen is where the cursor should be as illustrated first by Bolt [11]. It is the question *how* the ray is cast; directly following the extended index finger, as a ray between the eyes and the tip of the finger or as a ray from the wrist through the tip of the index finger [28]. The third of these three forms was used in our videoclip, see Figure 5.2a. In Bolt's set-up, users fully extended their arm and index finger for pointing;
2. *Repetitive taps*: making repetitive, discrete taps in the direction where the cursor should go as if using a keyboard's arrow keys. This is one of the gestures that we observed in the Wizard of Oz experiment and that was previously observed by Quek [172], see Figure 5.2b;
3. *Tap once*: the actor makes one tap and, by holding the hand in the direction where the cursor should go, the cursor will keep moving until the gesture is stopped explicitly by changing the hand shape to rest. This is a continuous alternative to the *Repetitive taps* method, see Figure 5.2c.

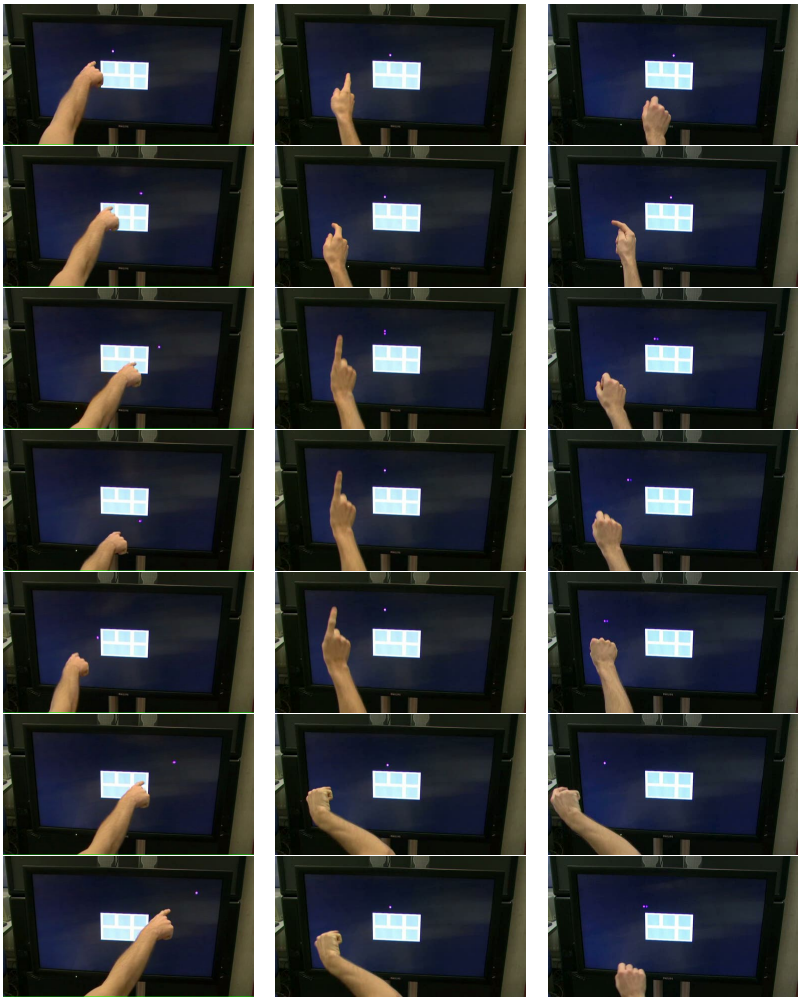


Figure 5.2: Gestures for pointing: (a) *Ray-casting*, (b) *Repetitive taps* and (c) *Tap once*.

Selecting

When a user is pointing towards some target, indirectly with a cursor or directly through ray-casting, the user should be able to select that object. The state transition in our four-state model is from tracking (#1) to selected (#2). The equivalent in a WIMP interface would be a left-button click action on the mouse: a mouse-down event followed by a button-up event [19]. Selected gestures are:

1. *AirTap*: tapping a target with the index finger as if pressing a mouse button in midair, named '*AirTap*' by Vogel and Balakrishnan [215]. Note that although this gesture mimics the exact motion users would use to operate a traditional mouse, there is no tactile feedback that confirms the action [156]. The *AirTap* gesture is depicted in Figure 5.6a;
2. *ThumbTrigger*: the hand is shaped like holding a pistol. The index finger and thumb are extended and by tapping the thumb on the index finger, or possibly the middle finger [64], the select action is performed, see Figure 5.3a. In contrast to *AirTap*, the user now receives tactile feedback while gesturing, caused by the thumb touching another finger;
3. *Dwelling*: the cursor dwells on a target for a brief time as introduced by [19; 241], see Figure 5.3b. This requires the hand to be kept still, for ray-casting the arm needs to be extended as well. Humans cannot keep their arm extended very well without slightly it trembling [114]. This becomes burdensome for the arms with fatigue setting in quickly;
4. *Encircling*: drawing a circle-shape with the cursor around a target after which the object is selected as is depicted in Figure 5.3c. Although we show a brief trailing motion of the cursor in our abstract application, it will be hard to predict the hand movements that show when the actor actually wants to select an object [26; 122; 233];
5. *FistGrab*: by closing the hand to a fist, as if grabbing something, the object where the cursor is located will become selected. It should be clear that this metaphor draws from everyday life where humans pick up objects in this manner. Note that in order to make a fist, the portrayed gesture requires all fingers to be stretched somewhat, see Figure 5.3d.

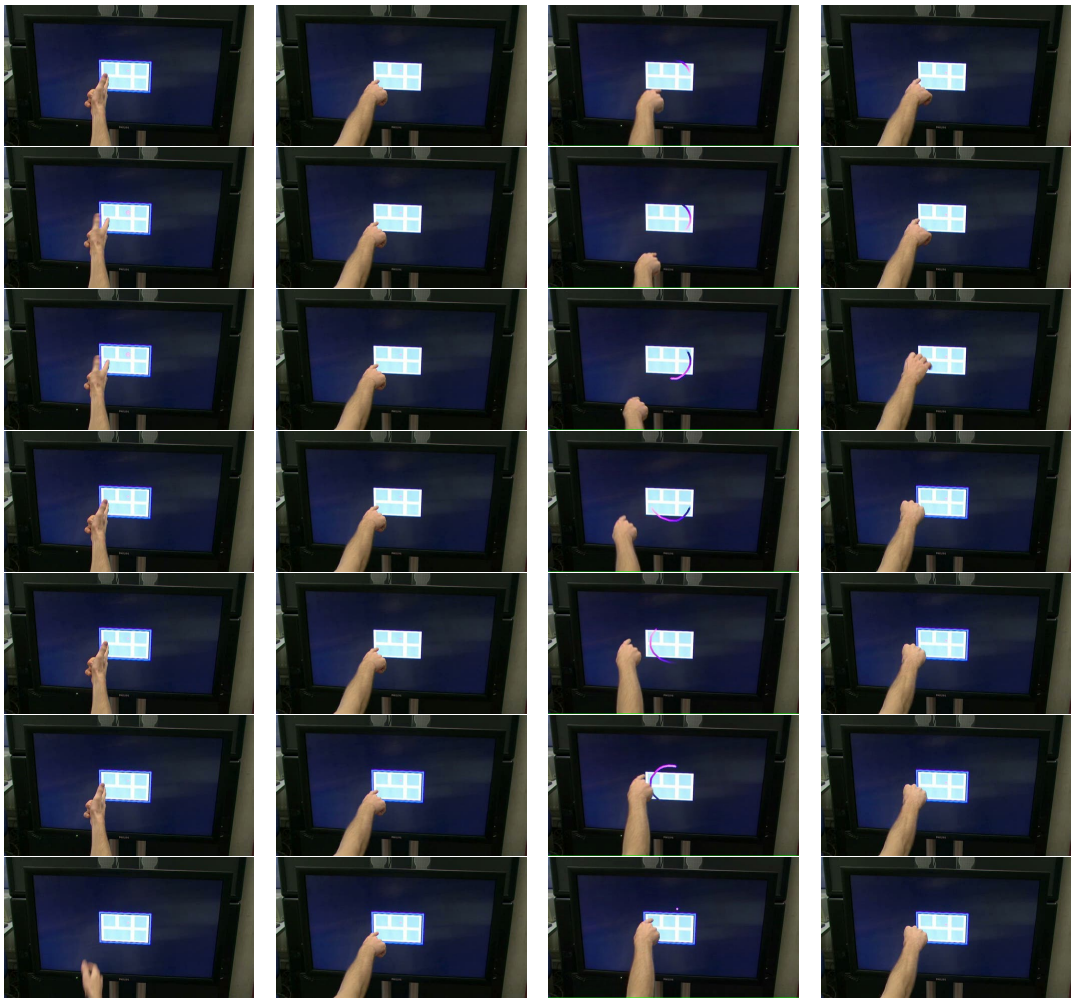


Figure 5.3: Gestures for selecting: *AirTap* [215] is depicted in Figure 5.6a, (a) *ThumbTrigger* [64], (b) *Dwelling* [241], (c) *Encircling* and (d) *FistGrab*.

Deselecting

When we are in the selected-state (#2), see Section 2.3, there should be a way to transition to the tracking-state (#1). In a mouse-based interface, this would be done inherently when clicking the mouse button: the button-down event switches to the selected-state while button-up directly changes back to tracking [19]. For this scenario we assume that the selection has already been done by means of one of the gestures that we described in Section 5.2. The gestures that we chose for deselecting are:

1. *DropIt*: opening the hand with the palm downwards as if dropping the selected target on the floor, see Figure 5.4a. Note that this is the direct opposite of the *FistGrab* gesture for selecting and that *DropIt* is thus also based on everyday life manipulation of real objects. The metaphor that we aimed for with this gesture is to drop a physical object on the floor;
2. *Retract to rest*: retracting the hand from pointing or gesturing to the display, placing it explicitly alongside the body as is depicted in Figure 5.4b. We have observed this gesture in the Wizard of Oz experiment, see Chapter 4. Note that by dropping the arm along side her body, the user is incapable of resuming the interaction right away;
3. *Jerky retract*: similar to the *Retract to rest* gesture only in a more condensed form. While locked onto an object, the hand is retracted briefly and in a jerky fashion. Note that *Jerky retract*, unlike *Retract to rest*, does not position the hand to rest alongside the body but that the arm remains extended towards the display, see Figure 5.4c. By keeping the arm extended the user is able to resume gesturing right away as opposed to having to lift her arm first;
4. *Select other*: selection of another target or selection of a blank space on the display using one of the selection methods that we mentioned in Section 5.2. This mimics the act of deselecting an icon or window in Windows-style operating systems. The videoclip depicted the *AirTap* gesture, see Figure 5.4d.

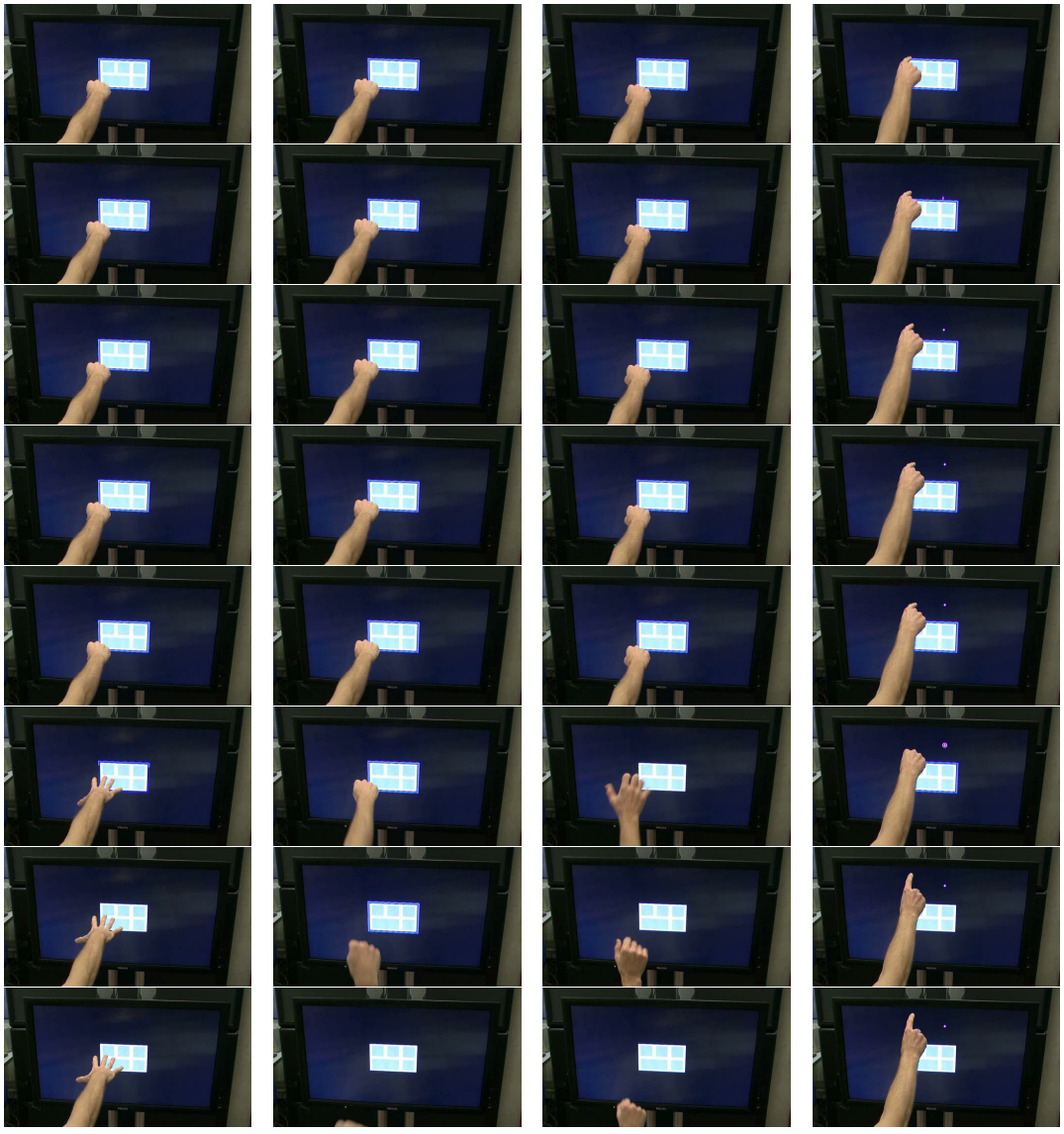


Figure 5.4: Gestures for deselecting: (a) *DropIt*, (b) *Retract to rest*, (c) *Jerky retract* and (d) *Select other*.

Resizing: shrinking and enlarging

Most demonstrator applications, especially for multi-touch interfaces [73], show an interface in which photos [237] or a map [189] are manipulated. In such interfaces, navigation consists of repeated tracking, selection and deselection events that accomplish positioning. Resizing is, much like orienting, a command that can be issued with a specific bimanual gesture when both hands are available for gesturing [135]. Note that the target is already selected by one of the gestures that were described in Section 5.2. Zooming in and out is similar to our description of resizing and it has been shown that users consistently made identical gestures for both commands [230]. We follow Wexelblat [226] who proposes to use the same gesture to issue two different commands depending on the application context. Enlarging and shrinking might be seen as two distinct commands, following selecting and deselecting. However, we will not present both alternatives separately to the user as they are parametrized in the gestures that we selected. These gestures also make resizing more predictable. Resizing is one implementation of the manipulation-state (#3). In our Wizard of Oz experiment we observed that participants mixed up enlarging and shrinking sometimes due to the limited amount of visual feedback from the interface. We expect that participants will automatically adjust their movements depending on the feedback of a functional interface. Gestures for shrinking and enlarging are:

1. *Fingers apart*: by moving two fingers of one hand apart to enlarge and then moving them towards each other to shrink is popular due to the Apple iPhone, see Figure 5.5a. Note that this gesture is rather small and that, by mapping it in an absolute way to a large display, will resize objects really fast. The distance between the fingers indicates the amount of resizing;
2. *Hands apart*: analogue to *Fingers apart* but on a larger scale. Instead of moving two fingers apart, in *Hands apart* both hands are moved apart enlarges and towards each other shrinks the object [135], see Figure 5.5b. The distance between the hands indicates the amount of resizing;
3. *PullPush*: grabbing with one hand and pulling the display contents closer, enlarging it in the process. Pulling the target close to enlarge it and pushing the target away to shrink it, as was observed in the Wizard of Oz experiment (refer to Chapter 4) and depicted in Figure 5.5c. In the portrayed gesture, we mapped the user's chest and his extended arm to fully zoomed in and fully zoomed out respectively;
4. *Referenced PullPush*[2]: the non-dominant hand represents a stationary point in the system while the dominant hand moves relative to the non-dominant hand [65]. Moving the hand in depth will either enlarge or shrink the target, see Figure 5.5d.

[2]A gesture for large display control as seen in Minority Report: see `http://www.imdb.com/title/tt0181689`, October 10th, 2009.

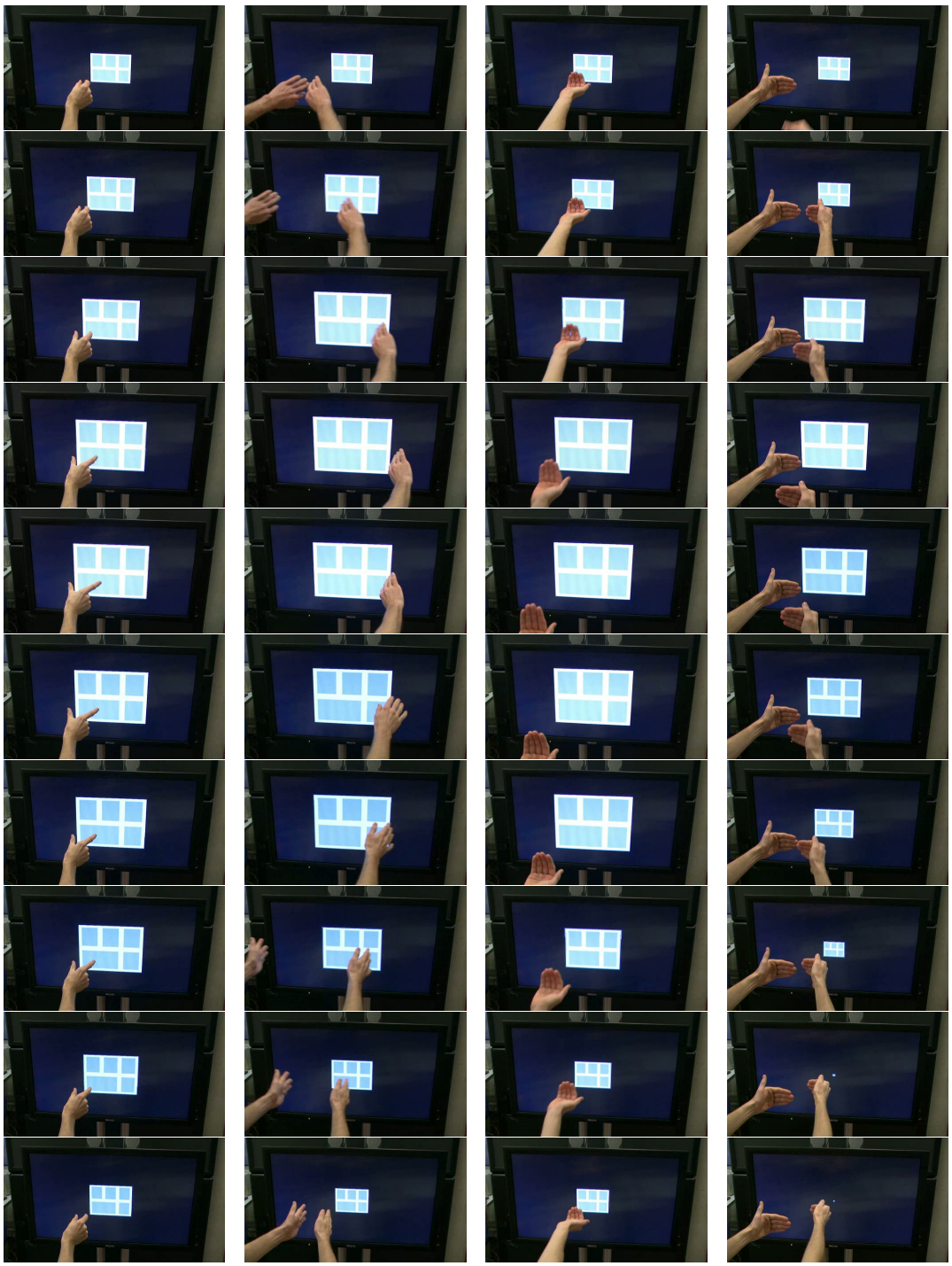


Figure 5.5: Gestures for resizing: (a) *Fingers apart*, (b) *Hands apart*, (c) *PullPush* and (d) *Referenced PullPush*.

Activate and deactivate

After a target has been selected, it should be activated in some manner. Like resizing, this is an implementation of the manipulation-state (#3). Clearly, this command, like resizing a target, is dependent on the semantics of the selected target. Activation can, for example, be possible if the selected target is an application or perhaps contains some information in the form of a hierarchy. Note that the selection is already done, depending on the gesture chosen for activation. Also like enlarging and shrinking, direct opposites of some of these gestures might also do the job. We selected them here based on their metaphorical application in everyday life. Gestures for this command are:

1. *AirTap*: double-click version of *AirTap* for selecting. The time between two clicks is minimal. Note that after one tap, the target becomes selected and that deactivation follows after a third *AirTap* gesture, see Figure 5.6a;
2. *AirTap & exit cross*: identical to the double-click *AirTap* for activating but deactivating is now done with a third *AirTap* on the exit cross of the window, see Figure 5.6b;
3. *ThumbTrigger*: double-click version of *ThumbTrigger* for selecting, following Grossman *et al.* [64]. In the videoclip, the thumb taps on the index finger. Note that after one tap, the target becomes selected and that deactivation follows after a third *ThumbTrigger* gesture, see Figure 5.6c;
4. *Dwell & exit cross*: identical to the *Dwelling* gesture for selecting. By keeping the cursor on the target for some additional time it will be activated [241]. Deactivating follows from performing *Dwelling* on an exit cross at the top of the object, see Figure 5.6d;
5. *Jerky PullPush*: pulling the selected target towards the actor in a short, jerky fashion by turning the hand quickly, see Figure 5.6e. Note that the arm moves from extended towards the body when the hand turns to activate the target and back to extended when deactivating;
6. *Open palm facing*: gently turning a flat hand with the flat palm towards the actor to activate the selected target as implemented by Vogel and Balakrishnan [214], see Figure 5.6f. To deactivate, the flat hand is turned back towards the display. Note that the arm remains extended;
7. *Drawing 'play' and 'stop' shapes*: similar to drawing a circular shape to select, the actor draws a 'play' shape (triangle) with the cursor to activate the target. To deactivate the target, a 'stop' shape (rectangle) is drawn. The play and stop shapes are analogue to the triangle and the square button, respectively, in audiovisual equipment, see Figure 5.6g. Note that the target is already selected before it can be activated;
8. *Activation and deactivation zones*: dragging a selected target to an 'activation zone' for activation, for example the left [214] or bottom [230] of the screen. Moving an activated target to the deactivation zone will deactivate the target, see Figure 5.6h.

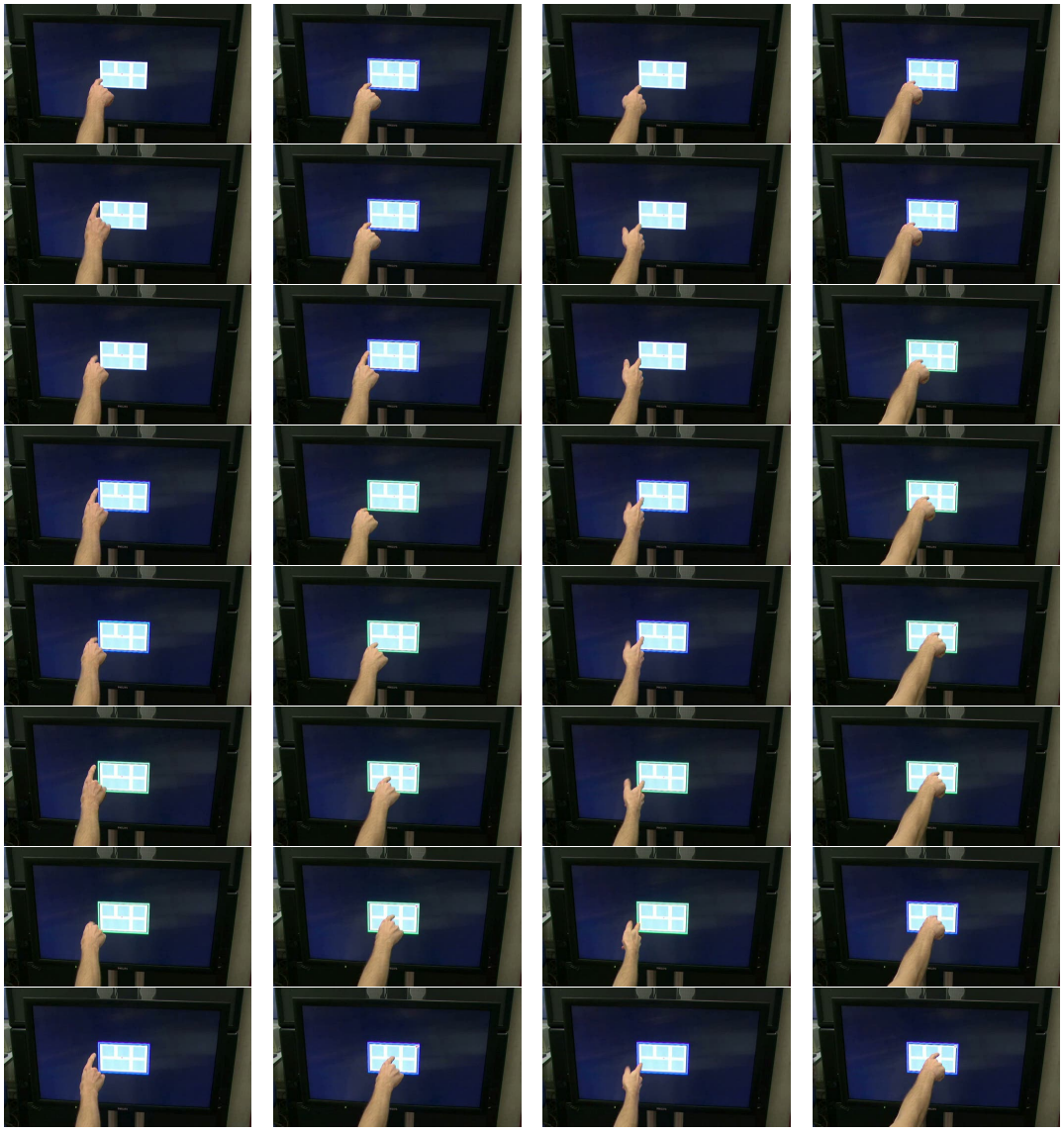


Figure 5.6: Gestures for activation and deactivation: (a) *AirTap* [215], (b) *AirTap & exit cross* [215], (c) *ThumbTrigger* [64] and (d) *Dwelling* [241]

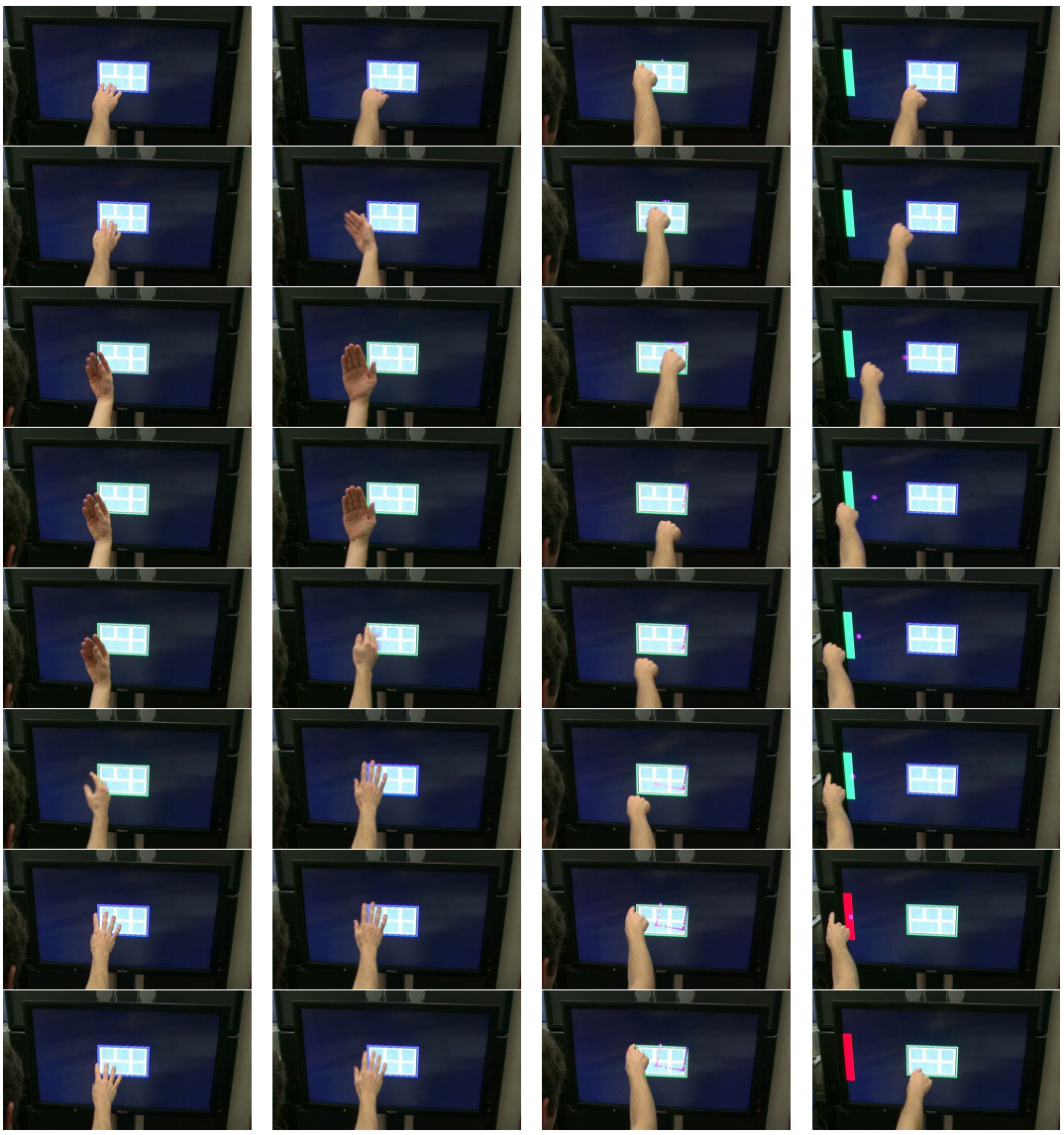
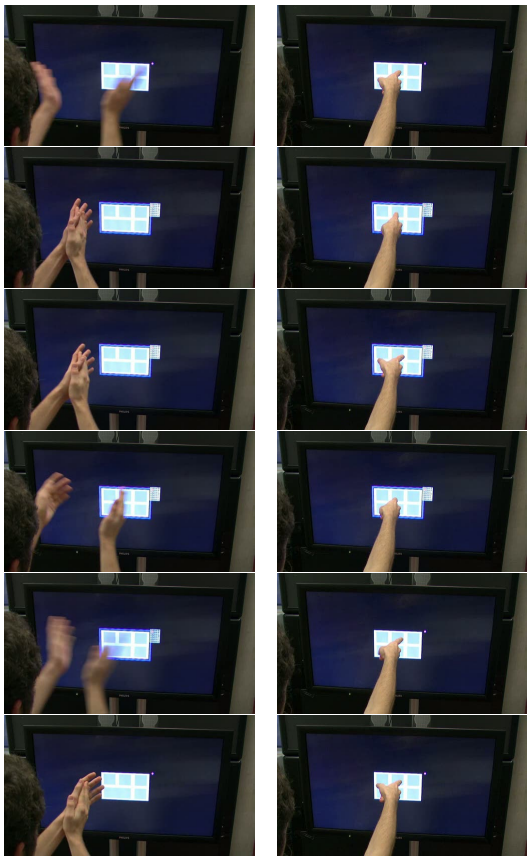


Figure 5.6: Gestures for activation and deactivation: (e) *Jerky PullPush*, (f) *Open palm facing* [214], (g) *Drawing 'play' and 'stop' shapes* (only 'stop' is shown here) and (h) *Activation and deactivation zones* (only activation is shown here) [214].

Context menu

One powerful addition in most desktop interfaces is the use of the right-button click on the mouse to open a menu or form of help in the current context. It is a third implementation of the manipulation-state (#3). This action is highly context-dependent so the cursor should be positioned first on a target, possibly blank space, where the menu should be opened. Target-selection might be an intermediate step yet in most current operating systems this is done as an implicit step. Gesture possibilities:

1. *Clapping*: clapping the hands to open the menu at the location where the actor pointed last while clapping them a second time will close the menu, see Figure 5.7a. Note that the hands interact with each other when clapping so that a cursor might be displaced in the process: this was not depicted in the video;
2. *PinkieTrigger*: similar to *ThumbTrigger* but the actor taps the thumb on the pinkie finger instead of on the index finger. The pinkie finger represents the right-button on a mouse. This gesture expands the *ThumbTrigger* gesture by Grossman *et al.* [64] for selecting. See Figure 5.7b.



(a) (b)

Figure 5.7: Gestures for opening and closing a context menu: (a) *Clapping* and (b) *PinkieTrigger*.

5.3 Results

In this section we report on the results of our online questionnaire. First, we describe our sample in Section 5.3.1. Second, we provide the ratings of gestures for each of the commands in Section 5.3.2.

5.3.1 Sample

A total of 99 subjects from five countries, the Netherlands, Germany, Belgium, the United Kingdom and the United States of America, participated in this within-subjects design. Participants were 28 years old on average (ranging 20-60 years, $\sigma = 8$ years). The questionnaire was completely filled out in roughly 25 minutes ($\sigma = 16$ minutes, ranging from 7 minutes to 91 minutes). We removed all incomplete trials (9 trials). As can be seen in Figure 5.8a, 22 subjects were female and 77 were male. In our sample, 25 subjects held a BSc's degree, 53 held a Master's degree and 12 held a PhD degree and 9 had another degree, see Figure 5.8b. The latter category consists of undergraduate students who have completed the Dutch high school (HAVO, VWO or MBO).

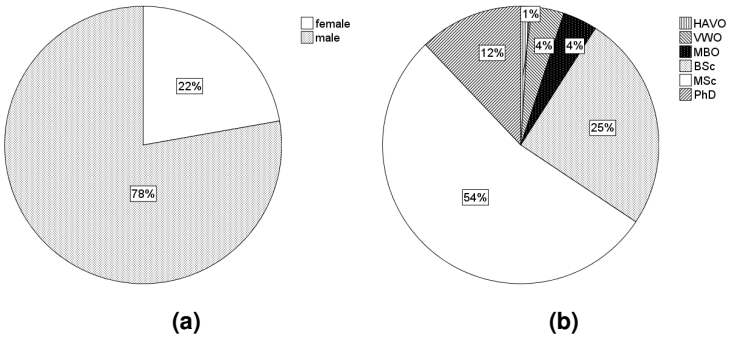


Figure 5.8: Sample characteristics (N = 99): (a) Ratio male-female and (b) education levels.

Our subjects were very knowledgeable of (online) videoclips in which gesture interfaces play a role ($\mu = 5.0$, $\sigma = 1.7$, with '1: unfamiliar (never seen one)' and '7: very familiar (regularly)'). They were proficient with the Apple iPhone ($\mu = 4.5$, $\sigma = 1.9$). They were moderately proficient with PDAs, smartphones or other pen-based hand-held devices ($\mu = 3.7$, $\sigma = 1.6$). Subjects were moderately proficient with other gesture interfaces ($\mu = 3.8$, $\sigma = 1.6$) on which most subjects remarked that they referred to devices such as the Nintendo Wii (11 subjects), mouse gestures in browsers (17 subjects), tablet PCs (6 subjects) and (multitouch) tabletop interfaces (7 subjects). For these latter three ratings (Apple iPhone, PDA or smartphone and gesture interfaces) we used the extremes '1: unfamiliar (what is that)' and '7: very familiar (own one)'. Using D'Agostino-Pearson K^2 analyses, we found normal distributions (with $p >= .05$) for our participants' familiarity with the Apple iPhone ($K^2 = 4.567$, $p = .10$), with PDA and smartphones ($K^2 = 3.230$, $p = .19$) and other

types of gesture interfaces ($K^2 = 4.725$, $p = .09$). The distribution was non-normal for our participants' familiarity with online videoclips ($K^2 = 8.588$, $p = .01$).

The results of our analyses for normality on the collected trials show that we cannot assume a normal distribution of our data. Intuitiveness scored $K^2 = 73.350$ ($p < .01$), physical effort $K^2 = 82.311$ ($p < .01$) and 'would use' scored $K^2 = 61.406$ ($p < .01$). This is, as is often the case in count-based data, caused by a ceiling or floor effect as is illustrated in Figure 5.9. Our trials data is further described in Table 5.1 where skewness and kurtosis are reported; the trials data are mostly deformed by a high variance which is shown by the low value for kurtosis. The Poisson-like distribution of our data cannot be transformed to a normal distribution through a logarithmic or square-root scale.

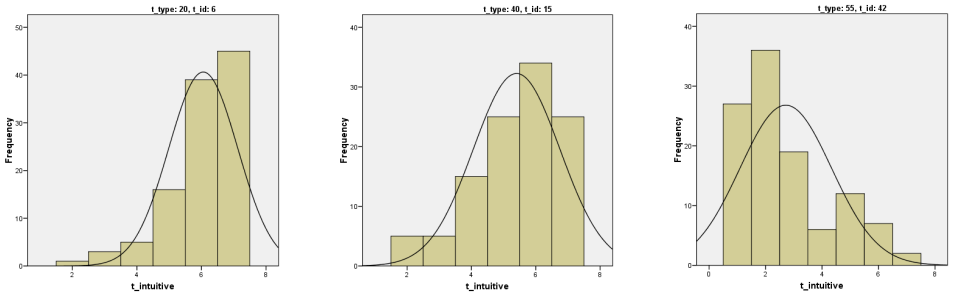


Figure 5.9: Histograms of our participants' scores on intuitiveness for commands 6, 15 and 42 respectively.

	intuitiveness	physical effort	'would use'
N	2574	2574	2574
mean	4.44	3.60	3.89
std. deviation	1.701	1.626	1.811
variance	2.892	2.643	3.280
kurtosis	-0.979	-0.769	-1.135
skewness	-0.238	0.340	0.039

Table 5.1: Description of the trials data from the online questionnaire.

5.3.2 Commands

Pointing

We found significant differences between gestures for intuitiveness ($\chi^2 = 106.098$, $p < .01$), physical effort ($\chi^2 = 61.827$, $p < .01$) and whether the participant would use this gesture ($\chi^2 = 138.275$, $p < .01$). *Ray-casting* scored significantly higher on intuitiveness and on 'would use' than both *Repetitive taps* and *Tap once*. There was no significant difference in scores on intuitiveness and on 'would use' between *Repetitive taps* and *Tap once*. *Repetitive taps* scored significantly lower than *Tap once* on 'would use'. Regarding physical effort, we found that *Ray-casting* scored best, significantly, followed by *Tap once* and then by *Repetitive taps*. Fourteen subjects

commented that *Ray-casting* is highly intuitive but that it would be fatiguing in the long run. Five subjects wondered how to stop pointing and proposed to stop when pointing off-screen.

Comments that we gathered from our participants included, for *Ray-casting*, "Very similar to Wiimote", "Not clear how to stop", "Seems the most intuitive method", "Repetitive use could cause new types of injuries?", "This makes perfect sense", "The Wim Kok mouse[3]", "Reflects human pointing". Also, some participants indicated that they would rather use a less tensed hand (1 participant) and either support the hand on a horizontal surface (2) or have more bent arms to reduce tension in the arms and shoulders (5). One participant mentioned the need of a jitter reduction algorithm to overcome the hand trembling slightly while pointing. Two participants argued that "it may be interesting to switch the pointer off" because in some cases "you just want to point out something" and "it is impossible to remove your hand [from the interface]". Comments on *Repetitive taps* were "Woody Woodpecker", "for fine-positioning perhaps [...] makes little sense for broad strokes" (2). Two participants argued that this could work when "you move the cursor on a grid" and "jumping between hyperlinks". Participants commented on the *Tap once* gesture that it would "require some physical effort" (1), "easy to learn", "I would prefer an open hand" (1), "speed should be adjusted by hardness of the hand" (1), "hard to perform accurate movements" (3), "overshoot will be a problem" (4), but they generally found it unintuitive (6). We did find that multiple participants (9) had some trouble with the videotaped gesture that included the actor making a fist to stop moving the cursor. These participants found the 'stop' command more problematic than tap once.

Selecting

There was a significant difference between gestures for intuitiveness ($\chi^2 = 98.816$, $p < .01$), physical effort ($\chi^2 = 58.266$, $p < .01$) and whether the participant would use this gesture ($\chi^2 = 80.725$, $p < .01$). *AirTap* scored best, significantly, on intuitiveness, followed by *ThumbTrigger*. There was no significant difference between *Dwelling*, *Encircling* and *FistGrab* concerning intuitiveness. *AirTap* also scored best, significantly, on 'would use' followed by *ThumbTrigger* which only outranked *FistGrab*, significantly. *AirTap* and *Dwelling* were ranked similarly for the amount of physical effort required to gesture while *AirTap* scored significantly lower than *Dwelling*, *Encircling* and *FistGrab*. *Encircling* scored significantly higher than the other gestures regarding physical effort.

Several participants noted that *AirTap* was based on the mouse (6) while some of them mentioned that they disliked that fact (2). Other participants commented that they would rather tap forward instead of downward (2) while some participants worried that using the index finger for clicking would hamper *Ray-casting* (2). On *ThumbTrigger* our participants argued that it can cause "problems in keeping the

[3]Wim Kok was the Dutch prime minister from 1994 till 2002. He once, in a primary school classroom, picked up a mouse and aimed it at a computer display where he wanted the mouse cursor to be, like a TV remote control.

index finger still while tapping" (4). Some participants commented that they "don't like the gun metaphor" (2) although others would have preferred to "pull the trigger" in this gun analogy (2). Participants commented on *Dwelling* that it required strain in holding position (5), that unwanted selections might occur (7) and that it "slows down the work flow" (7). Others disliked performing an action by inaction (2) and some commented that this was familiar on the iPhone (1). *Encircling* was "too slow" (4), "hard to use with many targets close to each other" (4) and "suitable for multiple targets but too much effort for just one" (4). Some participants worried how to distinguish between moving and encircling (2). *FistGrab* was regarded as natural and intuitive; "it seems like grabbing something" (4).

Deselecting

We found a significant difference between gestures for intuitiveness ($\chi^2 = 47.743$, $p < .01$), physical effort ($\chi^2 = 22.817$, $p < .01$) and whether the participant would use this gesture ($\chi^2 = 51.914$, $p < .01$). *DropIt* and *Select other* scored significantly better on intuitiveness (higher), physical effort (lower) and 'would use' (higher) than both *Retract to rest* and *Jerky retract*. There was no significant score difference between *DropIt* and *Select other* apart from a significant higher score for *Jerky retract* concerning physical effort. Twelve subjects indicated that they found selecting another target very proficient with windows-based interfaces. One of these subjects commented that although it is familiar, he would prefer 'some' other gesture.

DropIt was found (6) to complement *FistGrab* for selection. One participant (1) wondered whether this gesture might be more intuitive if the target actually dropped on the floor. Even though some participants (3) mentioned that they did not like this gesture they added that this was easy to remember. For *Retract to rest* some participants wondered when the arm is extended far enough (4). Others found this gesture tiring (2). Several participants (4) worried that this would hamper the interaction by slowing it down with selection rapidly following deselection. Participants commented on *Jerky retract* that "it isn't burning" (1) and that it feels unnatural (2). Several (12) participants commented that they were familiar with the *Select other* gesture from WIMP interfaces. Others (2) remarked that this is too much effort.

Resizing: enlarging and shrinking

We found a significant difference between gestures for intuitiveness ($\chi^2 = 74.200$, $p < .01$), physical effort ($\chi^2 = 64.381$, $p < .01$) and whether the participant would use this gesture ($\chi^2 = 64.117$, $p < .01$). There was no significant score difference between *Fingers apart* and *Hands apart* on intuitiveness and 'would use'. However, with respect to physical effort, *Hands apart* scored significantly higher than *Fingers apart*. *Referenced PullPush* scored significantly poorer on intuitiveness, physical effort and 'would use' with respect to the other three gestures apart from an insignificant difference with *Hands apart* regarding physical effort. Our participants scored *PullPush* significantly higher on 'would use' compared to *Fingers apart* while scoring it significantly lower than *Hands apart*.

For *Fingers apart* our participants commented that it is suitable for small resizes but not for larger resize commands (7). Most of these participants (4) added the comment that the fingers limit the maximum scalability. Several participants mentioned the iPhone explicitly as the source for this gesture (3). Also, some participants (5) argued that two hands make more sense when dealing with large displays. For *Hands apart* some participants mentioned that they would rather use one hand (1) or that this gesture is an extension of moving the fingers of one hand apart (3). Again, the iPhone was mentioned explicitly (2). *PullPush* was considered to be a large movement that "might be relieving [...] now and then" (1). Other comments mentioned that the whole display should zoom and not just one window (1) while the opposite was mentioned as well: "[this gesture] is most useful when there are a large number of objects on the screen, and you want to enlarge one of them". *Referenced PullPush* was considered too complex (1) and requiring too much effort (2), having a good "baseline reference for resize" (1). One participant found that, for Hands apart, the movie clip did not present the gesture as expected: "the window does not seem to be selected".

Activate and deactivate

There was a significant difference between gestures for intuitiveness ($\chi^2 = 140.976$, $p < .01$), physical effort ($\chi^2 = 121.518$, $p < .01$) and whether the participant would use this gesture ($\chi^2 = 154.250$, $p < .01$). *AirTap & exit cross* scored best, significantly, on intuitiveness and 'would use', followed by *AirTap*. *AirTap & exit cross* also scored significantly lower on physical effort than the other gestures while *AirTap* was not ranked significantly better than the remaining six gestures therein. *Drawing 'play' and 'stop' shapes* and using *Activation and deactivation zones* scored significantly poorer on all accounts when compared to all other gestures with *Activation and deactivation zones* as the worst alternative. *ThumbTrigger*, *Dwell & exit cross*, *Jerky PullPush* and *Open palm facing* did not score significantly different concerning the three ranked topics. Overall, we can distinguish three groups of gestures: the best gestures are *AirTap* and *AirTap & exit cross*, the worst gestures are *Drawing 'play' and 'stop' shapes* and using *Activation and deactivation zones*. The other gestures score in between with no significant differences.

Eight subjects commented that it was confusing to use the *AirTap* to both activate and select. With respect to *AirTap*, our participants commented "[do not] like activating to be the same movement as selecting" (1) while others mentioned this approach to be "ambiguous" (2). Extending *AirTap* to *AirTap & exit cross* gave these comments: "might be difficult to aim at the small exit cross" (2), "selecting and activating [should not] be the same action" (1), "[...] the cross is more like 'closing the window'" (1), but mostly participants argued that deactivation would be less handy (3). On *ThumbTrigger* our participants commented "again the cowboy gesture, I would go for it" (2), others again mentioned trying to replace the mouse with gestures (3) with terms such as "artificial". As with *AirTap & exit cross*, for *Dwell & exit cross* our participants found the exit cross too small (7) or that it resembled the Windows close button too much which was "the wrong icon" (3). *Jerky PullPush*

gave "I would mix up pull and push" (1) and "rotation of the wrist [...] causes strain" or "might cause repetitive strain injury" (5). Other participants worried that using depth as input will only work if depth is not used elsewhere. Comments on *Open palm facing* included "rotation of the wrist probably causes physical strain" (1), "my handicap [does not] allow me to make this particular gesture" (1) but mostly participants commented that keeping the hand open would be unpleasant (3). *Drawing 'play' and 'stop' shapes* comments were "costs more effort than classic control devices" (4) or that this gesture "take[s] too long" (2). Others wondered how more actions can be implemented in this manner, worrying that it might become too complex to remember properly (4). Concluding, comments on *Activation and deactivation zones* included "what does the red color mean?" (1), "euw" (1), "not very intuitive, nor very convenient" (1) and again "too much effort" (4). Other participants proposed to place objects along the perimeter with the (de)activation zone in the middle.

Context menu

A significant difference was found between gestures for intuitiveness ($\chi^2 = 1.993$, $p = .16$), physical effort ($\chi^2 = 28.795$, $p < .01$) and whether the participant would use this gesture ($\chi^2 = 7.098$, $p < .01$). We found a significantly better score for *PinkieTrigger* over *Clapping* for physical effort and 'would use'. However, the two gestures did not score significantly different on intuitiveness. Eleven subjects commented that both gestures feel rather awkward. Especially clapping would not be usable due to the noise it produces in public spaces. We found no significant influence on our two gestures for opening a context menu. For *Clapping* our participants commented that "one hand is better" (1) yet most were concerned that this gesture is "noisy" and dependent on the (public) space (8). Comments on *PinkieTrigger* included "not very fond of these high precision movements" (1), while other praised the low amount of effort required (2).

5.4 Summary of findings

We gathered subjective ratings on 26 gestures for six interface commands from a large, international sample (99 participants) with a similar background in an online questionnaire. For each of these six commands we found significant preferences for a specific gesture. Users expect a gesture-based interface to allow them to point directly at a target using pixel-precise ray-casting. For selecting, *AirTap* mimics clicking a mouse button very precisely even though no actual button can be pressed [215]. Some participants mentioned that they disliked the mouse metaphor but no participants mentioned the lack of physical feedback. Dwelling on a target is performing action through inaction which several participants explicitly disliked. Another participant proposed to 'throw away' an object to deselect or deactivate it. The gesture preferred for deselecting also leans heavily on existing interfaces where another target is selected to deselect the current target. The gesture used for

selecting (*AirTap*) was preferred to be used for activating and deactivating targets as well. Although some subjects indicated that they found it confusing to have the same gesture for two commands, we believe that this simplifies the interface which is indicated by the strong preference for this gesture. This also follows from the comments made by our subjects that they missed some way to 'click' for the resize command. For resizing, our subjects found that moving their fingers or hands apart was the most intuitive while some subjects wondered if moving two fingers apart would scale sufficiently to large displays. The gestures we offered for opening a menu were disliked, especially clapping the hands due to the noise it would make in public spaces.

5.5 Conclusions

Based on these findings we can construct a gesture set for explicit command-giving to a large display from beyond arm's length. The gesture set that was found in the experiment consists of: *Ray-casting* for pointing, *AirTap* and possibly *ThumbTrigger* for select and activating, *Fingers apart* and *Hands apart* as different scales for resizing. For opening a context menu, both gesture alternatives were disliked although *PinkieTrigger* scored best because it mimics *ThumbTrigger* which was the second-best gesture for select. The state-changing gestures that we have evaluated in this investigation are based on rigid hand shapes rather than on more complex, motion-dependent movements of the fingers, hands and arms. We have shown that in this approach that focuses on rigid hand shapes to explicitly signal state-transitions, the users of a gesture interface find the gestures intuitive and logical. Nielsen *et al.* [146] found that these qualities make the gestures easy to learn, remember and use in repeated use of the interface.

5.6 Discussion

The participants in this online questionnaire did not experience the gestures themselves in a working interface. This fact will most likely have influenced our findings but the extent is unknown until we compare these results with those that come from an actual working gesture interface. Chapter 6 describes our follow-up study that aims to validate the results that we found here. Another influence from the online questionnaire might be that we do not know if the videos worked correctly. It might have been so that users were prohibited from seeing some videoclips so that they could not fill out the questions based on that information. It might even have been that our participants filled out the questionnaire as quickly as possible without taking care to score meaningfully. The mean time that participants needed to fill out the questionnaire was 25 minutes and only three participants completed the investigation in 9 minutes or less so we assume that the participants took part seriously. We cannot be sure what the reasons are for sessions that lasted for more than 56 minutes (seven participants; equivalent to more than 2σ). However, we expect that

this can be explained by users taking a break or doing another chore in between.

In the videoclips we depicted a mock-up interface that was controlled by an operator behind the screens. One participant mentioned that the timing between the actor and the system's response was not always correct, for example, in ray-casting the cursor moved slightly ahead of the actor's arm. However, we do not expect that this will have influenced our results since our aim was not to depict a gesture interface as well as possible but to get an insight into the gestures that can be used to operate it. It is however, possible that the gestures that we recorded do not exactly match the literature or interface from which they stem. In addition, it might be that the clips did not show all details that define the gesture. For example, *ThumbTrigger* tapped the thumb on the extended index finger but it was not very clear how *Pinkie-Trigger* differed because the hand self-occluded the precise movements of the pinkie finger and thumb. This might have affected our results due to misunderstanding by our participants of the gesture that we proposed in the videoclip. However, we took care that the gestures were depicted precisely as described in Section 5.2 and their origins.

A common remark throughout the investigation was that the gestures should require as little effort as possible. This can also be seen in the scores as well: physical effort scored low when intuitiveness and 'would use' scored high consistently. This finding makes an argument for reusing the gestures as much as possible since that would make the interface more transparent, predictable and easy to use. Wexelblat [226] and Wu *et al.* [237] argue for reusing gestures depending on the context in which they are used for just these reasons. Our finding that *AirTap* is preferred throughout the interface is therefore not surprising.

Some participants proposed the use of other gestures. For example, using thumb-down for select, thumb-up for activate and closing the hand to deactivate. This specific example does not provide an out-of-range state (#0, see Section 2.3) where the hand is in rest. We argue that there should be some way for the user to rest his arm in order to prevent getting too fatigued. Some common remarks from our users focused on how to start and stop gesturing, for example, for pointing. This is rather similar to what our users are probably used to from their work with PCs where they position their hands from one device to the next when needed, for example, keyboard and mouse. In that setting it is possible to explicitly start interacting by touching the mouse or stop by letting it go [19]. This behaviour of positioning the hand or finger over a device or button is known as 'homing' and has been described in detail in the keystroke level model (KLM) by Card *et al.* [21]. As we found in our Wizard of Oz experiment, it seems that a way should be found to explicitly mark where a gesture begins and ends: when do we move to and from the out-of-range state (#0, see Section 2.3)? In our online questionnaire, we depicted the out-of-range state as pointing off-screen but we recognize that this topic might be more delicate. Wu *et al.* [237] define a gesture registration phase that is entered by a distinctive posture that, once recognized, sets the context for the dynamic and end phases. Various hand-shapes are employed to register a gesture, for example, index finger and thumb register writing while a fist registers the wipe gesture. This phase delineates one context from another to enable gesture reuse in various phases of the

entire compound gesture. This effectively starts and stops a gesture with explicit boundaries.

It is surprising in some way to discover that the familiarity of the mouse has such a strong impact on our findings. We readily accept that these results would be very different when consulting a user group that does not have this type of experience, from a different culture or even from another social group. However, the standard Windows-Icons-Menus-Pointing paradigm has, over the past decades, indoctrinated most users of the systems for which we are designing gesture-based interfaces. In that respect, we might even argue that the author of this thesis was born after the invention of the mouse and that he, like many other users of large display systems, grew up with this input device. We feel that the mouse has become such a strong metaphor that, even though a natural interface might be defined otherwise [158], it has by now become an everyday metaphor that drives the formation of a gesture set for explicit command-giving through hand gestures.

Chapter 6

Experiencing Gestures

> "Good judgement comes from experience. Unfortunately, experience comes from bad judgement."
>
> **Kenneth Boffard**
> [10, p.252] *Core Topics in General and Emergency Surgery*, vol. 1 of *Companion to specialist surgical practice series*, chap. 13, pp. 239-260. Elsevier Health Sciences, 3 ed.: 2006.

So far, we have explored how users perform gesturing in the absence of any instructions. Chapter 4 described the results of that exploration and we found that there is a great deal of similarity between users on the gestures that they spontaneously use to issue a command to the interface. This led us to believe that it might be possible to define a small set of gestures that can be used to issue commands in a HCI dialogue. In Chapter 5 we continued this line of research by investigating how a larger sample of similar background thinks about a large set of gestures that can be used for issuing a limited set of elementary commands to the interface. We found that there is a statistically significant consensus within a large user group on the gesture representations that should be used to achieve various goals in HCI dialogues.

We now relate our findings so far with our description of the interaction between human and computer in Section 2.1. First, we showed in Section 2.3 that there are some elementary interface tasks that a user can perform. Given a goal, say, finding a landmark on a map, the user then translates her goals to such elementary tasks. Second, we focused on the lexical level of performing a task, or, as we called it, issuing a command to the interface. We have shown that the fingers, hands and arms can serve as effectors with which the human user can control the interface. These effectors control the interface with movements that are idiosyncratic yet are considered intuitive by many users. The reason why these gestures are considered intuitive is that they minimize the mismatch between the psychological language that describes the user's goals on the one hand and, on the other hand, the action-oriented language of the interface that the user is trying to control. The transparent and easy to explain response of the interface is an integral part of it. We now understand why the gestures that we evaluated in the previous two chapters are considered intuitive and why others are not: intuitive gestures easily translate the

user's goals to tasks that can be issued through gesturing that, in turn, matches the interface's action-oriented language.

The results from the online questionnaire, see Chapter 5, that gave us access to a large user group might have been influenced by participants not fully understanding, appreciating and imagining what it would be like to issue commands in that way. It is hard to imagine and fully appreciate the workings of an interface without having experienced it. For one thing, it is difficult to imagine the lack of tactile feedback in an interaction with bare hands. This lack of feedback might even hamper those tasks where precision and feedback are crucial, for which applications that exploit multi-touch surfaces are prime examples [42]. In addition, any tactile feedback that is offered should be matched to other feedback, for example, visual feedback that the interface provides [156]. It is also hard to imagine what it would be like to perform a certain gesture repeatedly in an interface. Gestures might be strenuous for the hands and arms involved [215] or simply impossible to perform for certain users [146]. In this chapter, we will require subjects to perform gestures repeatedly, giving them the chance to experience rather than imagine the complete interaction. In doing so, we gain more insight into our findings so far. The interactions should last long enough for the user to fully appreciate the gesture and to comment on it.

This chapter is structured as follows. First, we describe in Section 6.1 the method used to validate the results from our online questionnaire. Second, Section 6.2 describes the results from this study and how they relate to the previous results. The findings of this investigation are summarized in Section 5.4. We draw conclusions in Section 6.4 and discuss these findings in Section 6.5.

6.1 Method of validating

To prevent biases from a different experiment set-up, we reused the set-up from the online questionnaire with some additions that we will describe in this section. Note that the commands and gestures that were described in Section 5.2 are integrally reused in this validation. The questionnaire itself was also reused to randomly present the gesture videoclips. The answers to questions on intuitiveness, the physical effort required and whether the subject would use that gesture for the given task on a seven-point Likert-style scale, see also Section 5.1, could be collected in this manner. For intuitiveness and 'would use', higher ratings translate to a better score while lower ratings for physical effort mean that the subject thinks that the gesture would require less effort to perform. In addition to these questions, we asked our participants to formulate a top-three of gestures for each task upon having performed all gestures. Participants were asked, after filling out all questions for a task, to comment on their preferences. They were also asked to provide gestures that they considered to be good alternative gestures.

For each gesture, after viewing its videoclip, we asked subjects to stand at a marked distance of two meters in front of a large display (52 inch diameter). This setting was the same as where we videotaped the video clips. There, they performed

the gesture at least three times while the abstract application, see Section 5.1.1, reacted to their hand(s) gesturing. Gesturing took place in a so-called gesture space that was directly in front of the participant, reaching to arm's-length [134, p.89]. The application was partially controlled by an operator who switched between the three application's states. The operator was introduced at the beginning of the investigation, participants were allowed to talk out loud to the operator. In order to get the participants to appreciate the gesture fully, the operator would ask them questions that addressed comfort and ease of use while they were gesturing. These questions aimed at engaging in a discussion on the gesture, not on judgement of it.

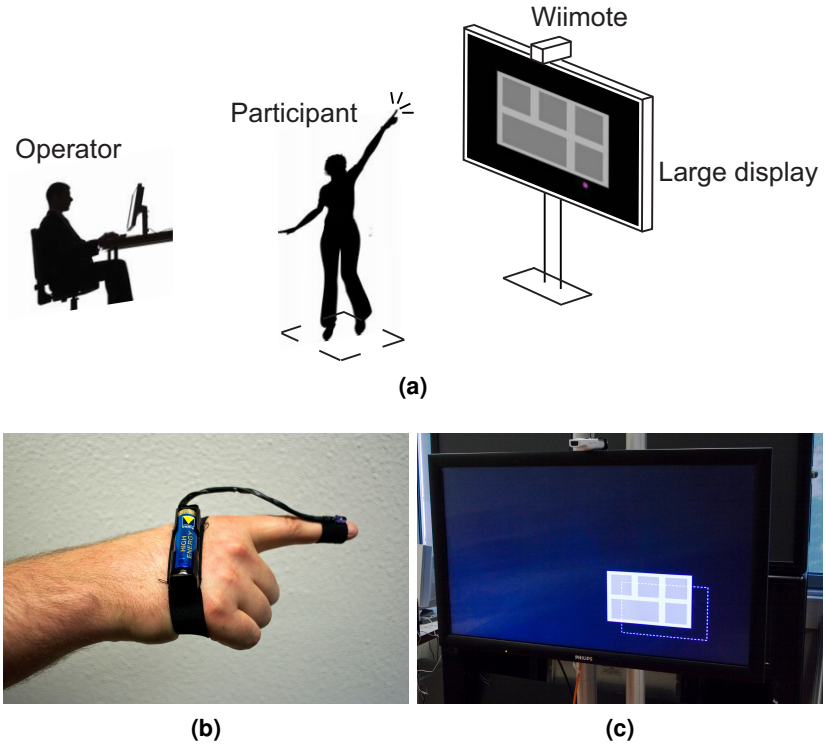


Figure 6.1: Experiment set-up for validating the findings of our online questionnaire: (a) a schematic overview with the position of operator, participant and the large display, (b) the 'glove' that our participants wore with an infrared LED mounted on the tip of the index finger and (c) the large display with a Wiimote mounted on top, facing the user.

Subjects wore a simple glove, see Figure 6.1b, that was made up from elastic bands to which an IR LED and AA battery were sewn. The experiment's set-up is depicted in Figure 6.1c. A Nintendo Wiimote[1] was mounted on top of the 52" large display. Its camera was used to detect the IR LED on the subject's index finger tip so that the cursor could be controlled through ray casting with an extended index finger [123]. The system was calibrated so that pointing, for example, to the top-left corner of the display would place the cursor there as well.

We extended the abstract application with some target locations that could be toggled on/off at various places on the display, an example of which is depicted

[1] `http://www.nintendo.com/wii`, September 26th, 2009.

in Figure 6.1c for dragging the on-screen object. The participants were asked to resize the application to the indicated target size or to point to the indicated target location. Concluding each trial, we asked each participant to point, select, drag and drop, resize, and activate the application given a series of target locations. We asked the participants afterwards if they felt in actual control of the application, as if the operator was not present. By doing so we wanted to get a feeling for how well our set-up succeeded in immersing the user in the interaction which we consider a measure for bias caused by the presence of the operator.

The validation that was described above was performed in two experiment conditions. First, we randomly selected ten percent of the participants in our online questionnaire based on availability. By comparing the ratings from this group with the ratings from the remaining 99 participants in the online questionnaire, we can assert whether the results from the questionnaire are a good representation. In other words, we can assert whether our participants could understand, appreciate and imagine how the gesture would work in an interface. Second, we asked a similarly sized group of volunteers with similar experience and who had *not* filled out the online questionnaire to participate in the same investigation. With this condition we investigate the potential bias that results from having filled out the online questionnaire and thus having seen the gestures before in the online videoclips. Although six months had passed between filling out the online questionnaire and the validation condition, we could not be sure that this would not influence our findings. In the remainder of this chapter we will denote each of the three conditions differently in order for the reader to recognize each condition with ease. The online questionnaire will be denoted as condition $Q1$. The validation condition in which users had already filled out the questionnaire entirely will be denoted condition Qx. The validation condition with only novice participants will be denoted condition Xp.

6.2 Results

In this section we report on the results of the validation conditions. First, we describe our samples in Section 6.2.1. Second, we provide the ratings of gestures for each of the commands in Section 6.2.2. The results reported in this section are achieved with a between-subjects design. A comparison between the three conditions is reported as well as an independent samples analysis of the trials data from conditions Qx and Xp. In addition, comments gathered from our subjects during the two conditions are summarized.

6.2.1 Sample

The two validation conditions were executed in an identical manner. The first user group, that of condition Qx, contained solely subjects who had already filled out the online questionnaire on an earlier moment. To ensure an independent samples analysis between these three experiment conditions we removed the participants in condition Qx from condition $Q1$. Note that the results reported in Chapter 5

only include participants from condition $Q1$: findings that were reported in Chapter 5 are thus independent from those reported in this chapter. For a description of the sample in $Q1$, see Section 5.3.1. The second user group, from condition Xp, consisted of subjects who had not previously filled out the online questionnaire.

Condition Qx

A total of ten subjects participated in this investigation, each one had filled out the online questionnaire at an earlier moment, roughly six months before. We selected the participants randomly and on their availability. Participants were 28 years old on average (ranging 24-36 years, $\sigma = 3$ years). The investigation took on average 53 minutes ($\sigma = 11$ minutes, ranging from 40 minutes to 70 minutes). All participants completed the investigation. One participant was female, the others were male. In our sample, one subject held a BSc's degree, seven held a Master's degree, one held a PhD degree and one was an undergraduate student. All participants were right-handed.

The participants were familiar with the Apple iPhone ($\mu = 4.0$, $\sigma = 1.6$), with PDAs and smartphones ($\mu = 5.3$, $\sigma = 1.7$) and with online videoclips ($\mu = 5.3$, $\sigma = 1.6$). In addition, they rated their familiarity with other gesture interfaces highly ($\mu = 4.9$, $\sigma = 2.1$). Examples of gesture interfaces that they meant by this are the Nintendo Wii and its Wiimote controller, prototypes developed at the university, touch-sensitive tables and other surfaces, data gloves and public ticket machines. A D'Agostino-Pearson K^2 analysis showed that there are normal distributions for the familiarity ratings with the iPhone, PDA and smart phones etcetera.

The trials data in condition Qx does not follow normal distributions: intuitiveness scored $K^2 = 17.545$ ($p < .01$), physical effort $K^2 = 6.786$ ($p = .03$) and 'would use' scored $K^2 = 7.683$ ($p = .02$). Our trials data is further described in Table 6.1 where skewness and kurtosis are reported; the trials data are mostly deformed as a result of low values for kurtosis.

	intuitiveness	physical effort	would use
N	260	260	260
mean	4.87	3.76	3.94
std. deviation	1.558	1.627	1.797
variance	2.426	2.648	3.228
kurtosis	-0.453	-1.019	-1.113
skewness	-0.633	0.077	-0.151

Table 6.1: Condition Qx: description of the trials data.

Condition Xp

Ten subjects participated in this investigation: none of whom had previously filled out the online questionnaire. The participants were 25 years old on average (ranging 22-29 years, $\sigma = 2$ years) and they needed 61 minutes to complete the investigation on average ($\sigma = 9$ minutes). All participants completed this condition of the experiment. Two participants were female, the others were male. In

this group, three subjects held a BSc's degree while the other seven held a Master's degree. One participant was left-handed, the other nine were right-handed. The participants were familiar with the Apple iPhone ($\mu = 5.3$, $\sigma = 1.4$) but not so much with PDAs and smartphones ($\mu = 3.4$, $\sigma = 1.4$). In addition, they were familiar with online videoclips ($\mu = 5.0$, $\sigma = 1.8$) and they rated their familiarity with other gesture interfaces highly ($\mu = 4.5$, $\sigma = 1.8$). The examples of such gesture interfaces that were mentioned were the Nintendo Wii and mouse gestures in the Opera browser. A D'Agostino-Pearson K^2 analysis shows that there is a normal distribution for participants personal answers regarding the familiarity with, for example, the Apple iPhone.

The data collected in condition Xp does not follow normal distributions: intuitiveness scored $K^2 = 18.600$ ($p < .01$), physical effort $K^2 = 11.039$ ($p < .01$) and 'would use' scored $K^2 = 11.549$ ($p < .01$). Our trials data is further described in Table 6.2 where skewness and kurtosis are reported; the trials data are mainly deformed as a result of low values for kurtosis and skewness.

	intuitiveness	physical effort	would use
N	260	260	260
mean	4.88	3.15	4.30
std. deviation	1.592	1.439	1.814
variance	2.535	2.071	3.292
kurtosis	-0.674	-0.439	-1.062
skewness	-0.619	0.470	-0.329

Table 6.2: Condition Xp: description of the trials data.

Sample summary

Comparing the three samples, we observed a significantly higher rating ($p = .02$) for the subject's familiarity with PDAs and smartphones in condition Qx compared to conditions $Q1$ and Xp. The other familiarity ratings did not differ between the three conditions.

Due to the non-normal distributions of our count-based data from conditions Qx and Xp we again make use of Kruskal-Wallis H analysis to discover differences between the ratings for the gestures per task. After finding significant differences we then examine those findings in more detail with pair-wise Mann-Whitney U analyses. We relate our findings to those reported in Chapter 5 that described the online questionnaire condition ($Q1$).

6.2.2 Commands

Pointing

In a comparison between the three conditions we found no significant differences between our three conditions on intuitiveness ($\chi^2 = .845$, $p = .66$), physical effort ($\chi^2 = 2.252$, $p = .32$) and 'would use' ($\chi^2 = 2.718$, $p = .26$). The analysis per gesture can be found in Table 6.3.

	condition	$Q1$	Qx	Xp		
	N	99	10	10	χ^2	p
Ray-casting	Intuitiveness	59.13	78.35	56.40	3.454	.18
	Physical effort	59.54	75.40	55.20	2.229	.33
	Would use it	61.92	48.25	58.60	1.552	.46
Repetitive taps	Intuitiveness	58.35	70.35	65.95	1.468	.48
	Physical effort	60.23	62.85	54.85	.316	.85
	Would use it	59.55	72.20	52.25	1.894	.39
Tap once	Intuitiveness	61.74	53.75	49.00	1.659	.44
	Physical effort	59.46	69.80	55.55	1.039	.60
	Would use it	63.43	42.25	43.80	6.097	$< .05$

Table 6.3: Differences between the online ($Q1$) condition and validation (Qx and Xp) conditions for the *point* gestures. Kruskal-Wallis H analyses results with mean ranks are reported. Insignificant results have been shaded.

In the trials data from condition Qx we found a significant difference between gestures for intuitiveness and whether the participant would use this gesture but we did not observe a significant difference for physical effort. Comparing the results from condition Qx to those found in condition $Q1$ we found identical though not so strongly pronounced ratings. A significant difference was found in the trials data from condition Xp for all three questions. Compared to the results in conditions $Q1$ and Qx these differences are similar. The top-three rankings for pointing gestures showed a clear preference (9 subjects) for *Ray-casting* over its alternatives. However, in the online questionnaire ($Q1$) our participants rated *Ray-casting* significantly lower on physical effort but that was not the case in conditions Qx and Xp. Our participants expected that a prolonged interaction causes fatigue in the arms for holding them outstretched while pointing.

In conditions Qx and Xp, our participants wondered whether fine movements for *Ray-casting* would not suffer from jitter. For *Repetitive taps* our subjects argued that it would be viable for small distances where precision is required but that it is very unsuitable for long distances, mainly due to fatigue. The same comments were made concerning the time spent in interaction: longer tasks would fatigue the user too much. One participant mentioned his preference for pointing with the whole hand instead of only the index finger. With respect to both tapping gestures, some subjects argued that it would work better when using both hands: when moving the cursor to the right the left hand would be better suited whereas the right hand is best for moving the cursor left. For *Tap once* our users found it hard to time when to stop the cursor movement. As an alternative pointing gesture, it was proposed to use some gesture, for example, *AirTap*, to switch the cursor between objects on the screen. In both conditions, participants mentioned their preference to combine these pointing gestures. For example, *Ray-casting* provides an easy means to cross large distances and fine-tuning can be accomplished with *Repetitive taps*. We observed in all three gestures that most users will bend their preferred hand in awkward poses so that they can keep pointing with their index finger.

Alternatively, *Tap once* could also be implemented with a deceleration measure so that the cursor could be 'thrown' across the surface until it would stop automatically.

The distance traveled is then based on the intensity of the gesture as has been implemented by the BumpTop interface [1]. In BumpTop, icons are represented as physical objects that behave in a believable manner due to a physics simulation. Another alternative gesture that was mentioned was *Ray-casting* with just finger movements while the hand and arm are left in rest, for example, alongside the body. This gesture would depend to large extent on the visual feedback on the display.

Selecting

Comparing the ratings for the select gestures in the three conditions we found no significant differences for intuitiveness ($\chi^2 = 2.912$, $p = .23$), physical effort ($\chi^2 = 4.104$, $p = .13$) and 'would use' ($\chi^2 = 4.572$, $p = .10$). However, when taking a more detailed look at the ratings per task in Table 6.4, we see a significantly higher rating in conditions Qx and Xp for *ThumbTrigger* on intuitiveness and 'would use' compared to conditions $Q1$ while the rating for physical effort scored significantly lower in conditions Qx and Xp than it did in conditions $Q1$.

	condition	$Q1$	Qx	Xp		
	N	99	10	10	χ^2	p
AirTap	Intuitiveness	61.86	40.25	61.35	4.029	.13
	Physical effort	60.83	62.25	49.55	1.093	.58
	Would use it	61.49	37.80	67.45	5.280	.07
ThumbTrigger	Intuitiveness	53.41	90.50	94.75	22.455	< .01
	Physical effort	63.59	39.85	44.65	6.796	< .05
	Would use it	54.37	83.75	92.00	16.417	< .01
Dwelling	Intuitiveness	62.35	49.35	47.35	2.921	.23
	Physical effort	58.36	81.60	54.65	4.569	.10
	Would use it	61.19	54.25	53.95	.725	.70
Encircling	Intuitiveness	59.51	68.35	56.55	.733	.69
	Physical effort	60.55	66.70	47.90	1.691	.43
	Would use it	60.22	56.55	61.30	.122	.94
FistGrab	Intuitiveness	57.90	70.15	70.60	2.272	.32
	Physical effort	59.84	68.90	52.70	1.178	.56
	Would use it	58.74	59.75	72.75	1.552	.46

Table 6.4: Differences between the online ($Q1$) and validation (Qx and Xp) conditions for the *select* gestures. Kruskal-Wallis H analyses results with mean ranks are reported. Insignificant results have been shaded.

The results from condition Qx show a significant difference between the gestures for intuitiveness, physical effort and whether the participant would use this gesture. The preference for a specific gesture is less pronounced due to the smaller user group in condition Qx. The overwhelming preference for *AirTap* in condition $Q1$ is not present in condition Qx: participants rated *AirTap*, *ThumbTrigger*, *Encircling* and *FistGrab* similarly with respect to intuitiveness. *AirTap* also scored similarly to *ThumbTrigger* and *FistGrab* based on physical effort but both *Dwelling* and *Encircling* scored significantly higher. *ThumbTrigger* did score higher in conditions Qx and Xp with respect to intuitiveness than *Dwelling*, *Encircling* and *FistGrab*. In addition, *ThumbTrigger* scored significantly lower in conditions Qx and Xp on physical effort

except when compared to *AirTap*. Participants would use either *AirTap*, *ThumbTrigger* or *FistGrab* to issue a select command where the difference between *AirTap* and *ThumbTrigger* with the other three gestures was the most pronounced. For the select gestures in condition Xp, a difference was revealed for intuitiveness and for physical effort but not for whether the participant would use this gesture. Comparing the findings from condition Xp to condition $Q1$ we see the same differences as are described above in the comparison between conditions Qx and $Q1$. However, the differences between the gestures are more pronounced in condition Xp than they are in condition Qx. In total, six subjects from condition Qx rated *ThumbTrigger* as the best gesture, closely followed by *AirTap* which was placed as second best by five subjects. In condition Xp, there was a draw of four subjects each preferring one of the gestures and five subjects placing *AirTap* or *ThumbTrigger* as second best. Third best in both conditions was *FistGrab*.

In comments in conditions Qx and Xp our subjects felt that *AirTap* was very familiar to the mouse-paradigm but that it would be preferred if you could tap towards the screen ("as if pressing a button in a lift") instead of having to press your finger down. In doing so, the cursor would also remain more stationary. *ThumbTrigger* also mimicked the mouse-paradigm while some participants compared it to a pistol-shaped hand. Our subjects liked the fact that *ThumbTrigger* allows selecting while pointing but they argued to relax the hand somewhat instead of keeping the middle, ring and pinkie fingers bent. In addition, some subjects mentioned that it is nice to separate the act of pointing from the act of selecting. *Dwelling* was considered to be inaccurate with possibilities for accidental selection events in addition to taking too long to select an object. *Encircling* took too much effort and time but was considered to be suitable for multiple-object selection. *FistGrab* was familiar from everyday life but was linked more to picking up and moving objects (dragging) than for selecting them. Some participants commented that both *FistGrab* and *ThumbTrigger* would move the index finger when changing the hand tension which reduced pointing accuracy. Others mentioned that they preferred to use *ThumbTrigger* when pinching the tips of the middle finger and the thumb together to relieve tension in the hand.

Deselecting

There were no significant differences between the three conditions for the ratings of the deselect gestures intuitiveness ($\chi^2 = 3.386$, $p = .18$), physical effort ($\chi^2 = .862$, $p = .65$) and 'would use' ($\chi^2 = 2.942$, $p = .23$). This is also illustrated in Table 6.5 where the analysis results are given per gesture. In condition Qx, we did not find a significant difference between gestures for intuitiveness and whether the participant would use this gesture. The difference for physical effort was significant however ($\chi^2 = 6.092$, $p < .05$). Our analysis results show that *DropIt* did not score significantly higher on intuitiveness than *Retract to rest* and *Jerky retract* which it did in condition $Q1$. With respect to physical effort and 'would use', we found the same results as in condition $Q1$ although the differences were less significant. Contrary to these findings for the trials data from condition Qx, the data from condition Xp revealed a difference on intuitiveness ($\chi^2 = 13.850$, $p < .01$) and whether the

participant would use this gesture ($\chi^2 = 14.995$, $p < .01$) but not for physical effort. The results from conditions Qx and Xp are largely the same, however, there is a significant difference between preference for *Select other* over *DropIt* with respect to intuitiveness. That difference was not observed in condition $Q1$. Subjects in conditions Qx and Xp placed *Select other* on top in their rankings closely followed by *DropIt* and *Jerky retract*.

	condition	$Q1$	Qx	Xp		
	N	99	10	10	χ^2	p
DropIt	Intuitiveness	58.84	69.00	62.50	.888	.64
	Physical effort	60.96	55.20	55.25	.492	.78
	Would use it	58.11	66.45	72.30	1.985	.37
Jerky retract	Intuitiveness	59.00	66.90	63.00	.583	.75
	Physical effort	59.82	69.75	52.00	1.390	.50
	Would use it	59.95	52.30	68.20	1.098	.58
Retract to rest	Intuitiveness	60.06	64.25	55.20	.359	.84
	Physical effort	58.74	67.45	65.05	.834	.66
	Would use it	59.49	67.95	57.10	.649	.72
Select other	Intuitiveness	59.89	70.15	80.65	5.605	.06
	Physical effort	61.53	54.55	50.30	1.314	.52
	Would use it	57.11	78.00	70.60	4.606	.10

Table 6.5: Differences between the online ($Q1$) and validation (Qx and Xp) conditions for the *deselect* gestures. Kruskal-Wallis H analyses results with mean ranks are reported. Insignificant results have been shaded.

Our subjects commented that *Select other* was very familiar from computer operating systems such as Windows and Mac OS X. They argued that deselect of individual targets should be possible when having selected multiple objects in a row. In addition, although we showed *AirTap* for selecting something other than the on-screen object, most subjects spontaneously used their preferred select-gesture, in most cases: *ThumbTrigger*. *DropIt* looked similar to the hand shape when *Raycasting* and our subjects wondered how this gesture is started when, for example, *AirTap* was used to select an object. In addition, the difference between the relaxed hand shape and *DropIt* was thought to be too subtle. On the other hand, some participants did mention that *DropIt* is the opposite of *FistGrab* and that these two gestures might be suitable for dragging an object instead of for (de)selecting. For *Retract to rest*, our subjects commented that the arm movements were too large when having to move back and forth between rest and the gesture space. In that respect, *Jerky retract* was better because it leaves the arm in the gesture space. However, the jerky movement strained the arm which was disliked. It was proposed by multiple subjects in conditions Qx and Xp that select and activate should be explictly split, with distinct gestures for each task. One proposed gesture was to use a variant of *AirTap* with the middle finger to activate while *AirTap* with the index finger would select.

Resizing: enlarging and shrinking

Comparing the three conditions on the gestures for resizing, we found no significant differences in intuitiveness ($\chi^2 = 3.936$, $p = .14$), physical effort ($\chi^2 = 2.726$, $p = .26$) and 'would use' ($\chi^2 = .398$, $p = .82$). We did find some differences when analysing the ratings of the three conditions per gesture, see Table 6.6. Although there was no difference in whether the participants would use it, *Referenced PullPush* scored significantly higher on intuitiveness and lower on the physical effort required to perform the gesture in conditions Qx and Xp. Similarly, participants scored significantly higher on 'would use' for *Hands apart* than in condition $Q1$.

	condition	$Q1$	Qx	Xp		
	N	99	10	10	χ^2	p
Fingers apart	Intuitiveness	61.36	60.80	45.75	1.982	.37
	Physical effort	58.78	59.80	72.30	1.522	.47
	Would use it	63.67	43.20	40.45	6.951	< .05
Hands apart	Intuitiveness	56.45	76.28	74.55	5.337	.07
	Physical effort	62.93	44.67	38.85	6.573	< .05
	Would use it	55.45	79.89	81.25	9.060	.01
PullPush	Intuitiveness	59.49	65.30	59.75	.272	.87
	Physical effort	58.91	63.50	67.30	.676	.71
	Would use it	60.67	57.35	56.05	.236	.89
Referenced PullPush	Intuitiveness	56.26	80.20	76.80	7.240	< .05
	Physical effort	63.56	54.25	30.50	8.978	.01
	Would use it	57.38	67.10	78.85	4.104	.13

Table 6.6: Differences between the online ($Q1$) and validation (Qx and Xp) conditions for the *resize* gestures. Kruskal-Wallis H analyses results with mean ranks are reported. Insignificant results have been shaded.

Looking at the results from condition Qx, we found no significant differences between the resize gestures for each of the three questions: intuitiveness, physical effort and 'would use'. We found that *Hands apart* scored better than the other three gestures although the difference with *Fingers apart* was barely significant. *Referenced PullPush* scored similarly to *Fingers apart* in condition Qx while in condition $Q1$ the latter was found far more intuitive. The preference found in condition $Q1$ with respect to physical effort were absent in condition Qx except for *Referenced PullPush* that had a significantly higher score compared to *Fingers apart*. Condition Xp did show a significant difference between the four resize gestures on intuitiveness ($\chi^2 = 11.322$, $p = .01$) and whether the participant would use this gesture ($\chi^2 = 11.801$, $p < .01$). Comparing the findings of conditions Qx and Xp we see largely the same results although the physical effort required to perform *Referenced PullPush* is not different from that of *Fingers apart*. In addition, *Hands apart* scored significantly better than *Fingers apart* on both intuitiveness and whether the participant would use the gesture. The subjects in both conditions ranked *Hands apart* as the best gesture for resizing. In condition Qx, *Fingers apart* was ranked second while *PullPush* was ranked second in condition Xp. For both conditions, *Referenced PullPush* was ranked third best.

For *Fingers apart*, our subjects argued that for minor changes in size, this gesture

would work as adequately as it would for small displays. However, for larger changes, the fingers would have to repeatedly gesture from start to stop which made the gesture physically taxing. Also, the starting-posture was a bit difficult to determine. Some subjects mentioned that they felt that this gesture is only suited for smaller displays due to the match in their physical sizes. Quek [172] described this repeated form of gesturing as beats or strokes as we have described them in Section 3.2. The *Hands apart* gesture was more precise in that respect although one subject would have preferred it if the distance between the hands had matched the object's size from the user perspective. For *PullPush*, it was also hard to determine the starting-posture and the limited arm-length introduced a problem for larger zoom-ranges. *Referenced PullPush* was very novel and our subjects liked the reference to the starting position although it could be more explicit, said one subject, by adding a 'click' sound when placing both hands together. However, having to use both hands was more tiring. Three subjects, all from condition Xp, tried to spontaneously move both hands in *Referenced PullPush* while the others had to be told explicitly. By moving the reference-hand, the zoom-range could be extended. For *Fingers apart*, one subject proposed the use of the hand's distance to the body as a means to accelerate or decelerate the resize speed. For both *Fingers apart* and *PullPush*, our subjects had difficulty traversing larger resize-ranges. They mentioned to 'have to pick up the mouse' and to 'need to re-gesture' with which they meant that the same gesture had to performed repeatedly while moving the hand from and to the gesture space in between repetitions.

Activate and deactivate

We found significant differences between the three conditions for the activate and deactivate gestures on intuitiveness ($\chi^2 = 22.832$, $p < .01$), physical effort ($\chi^2 = 15.373$, $p < .01$) and 'would use' ($\chi^2 = 7.409$, $p < .05$). Similar to our findings for the select task, *ThumbTrigger* scored significantly higher on intuitiveness and whether the participants would use that gesture in conditions Qx and Xp compared to condition $Q1$. In addition, it scored significantly lower on the amount of physical effort required to gesture, see Table 6.7. The results for *Dwell & exit cross* were diverse: from condition Qx we observe that a higher effort is required than in condition $Q1$ while condition Xp reveals the opposite. In both conditions Qx and Xp, the participants scored significantly lower on whether they would use this gesture. For *Jerky PullPush*, we found that the subjects in condition Xp scored intuitiveness and 'would use' significantly higher while the required effort scored significantly lower. *Open palm facing* scored significantly higher on intuitiveness in conditions Qx and Xp and although our subjects did score higher in those conditions for 'would use' than they did in condition Qx, the difference was not significant. In both conditions, our subjects ranked *ThumbTrigger* as the best gesture. *AirTap* was ranked second best in condition Xp and third best in condition Qx. The subjects in condition Qx ranked *Open palm facing* as second best, while *Jerky PullPush* was ranked third best in condition Xp.

Our participants found both *Dwell & exit cross* and *AirTap & exit cross* familiar

	condition	$Q1$	Qx	Xp		
	N	99	10	10	χ^2	p
AirTap	Intuitiveness	58.40	59.55	76.25	2.577	.28
	Physical effort	60.19	67.40	50.75	1.287	.53
	Would use it	59.64	54.45	69.15	1.021	.60
AirTap & exit cross	Intuitiveness	58.87	54.30	76.85	2.897	.24
	Physical effort	59.87	75.80	45.45	4.095	.13
	Would use it	60.68	38.80	74.50	5.759	.06
ThumbTrigger	Intuitiveness	53.53	90.65	93.45	21.468	< .01
	Physical effort	64.06	44.75	35.10	8.942	.01
	Would use it	53.23	88.45	98.60	23.788	< .01
Dwell & exit cross	Intuitiveness	59.80	66.40	55.55	.532	.77
	Physical effort	58.40	91.35	44.45	10.939	< .01
	Would use it	64.16	40.85	38.00	8.922	.01
Jerky PullPush	Intuitiveness	56.09	71.55	87.15	8.900	< .05
	Physical effort	62.38	65.50	30.95	8.271	< .05
	Would use it	57.22	60.10	87.40	7.177	< .05
Open palm facing	Intuitiveness	56.25	84.85	72.30	7.884	< .05
	Physical effort	60.41	70.30	45.60	2.775	.25
	Would use it	56.63	78.85	74.55	5.907	.05
Drawing 'play' and 'stop' shapes	Intuitiveness	57.35	78.00	68.20	4.093	.13
	Physical effort	61.96	49.60	51.00	1.988	.37
	Would use it	59.61	60.40	63.50	.127	.94
Activation and deactivation zones	Intuitiveness	58.81	63.15	68.65	.866	.65
	Physical effort	60.17	64.70	53.60	.550	.76
	Would use it	59.87	59.75	61.50	.022	.99

Table 6.7: Differences between the online ($Q1$) and validation (Qx and Xp) conditions for the *(de)activate* gestures. Kruskal-Wallis H analyses results with mean ranks are reported. Insignificant results have been shaded.

from Windows-based operating systems. One participant proposed the combination of *ThumbTrigger* with such an exit cross. Others found it difficult to precisely point at the exit cross due to jitter from *Ray-casting* and it was frequently mentioned that this Windows-metaphor should be abolished in such novel interfaces: 'better solutions should be found'. *AirTap* and *ThumbTrigger* were preferred for their speed in switching states although our subjects wondered how toggling between point, select and activate would work in an actual system. Like selecting, *Dwell & exit cross* required too much time to switch between states. *Jerky PullPush* was less stressful for the hand compared to having to stretch it as in *Open palm facing*. However, users felt that it was hard to determine for *Jerky PullPush* when the gesture was 'good enough' to be recognized. Our participants found *Activation and deactivation zones* too bothersome as these required them to perform an additional dragging task to (de)activate the object. It was proposed to move this zone closer to the object to minimize this additional task, for example, attaching an activation zone to the left and a deactivation zone to the right of the selected object. Similarly, the *Drawing 'play' and 'stop' shapes* gesture required the participants to draw error prone shapes. One subject proposed the use of more simple shapes, for example, those that are inherent in the browser Opera, see Figure 6.2. It was argued that a gesture that was distinctly different from select and deselect would be a more explicit interaction

signal for both user and system. In that respect, some subjects argued that double clicking, for example, using *AirTap* or *ThumbTrigger*, would be a better gesture.

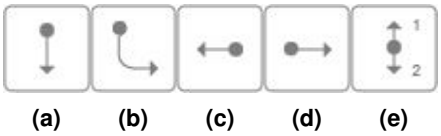


Figure 6.2: Mouse gestures in the Opera browser[2] (a) Open new document, (b) close document, (c) previous page in history, (d) next page in history and (e) reload document.

Context menu

In a comparison between the three conditions for the context menu gestures, we found a significant difference on whether the participants would use the gesture ($\chi^2 = 9.478$, $p < .01$) but not for intuitiveness ($\chi^2 = 5.414$, $p = .07$) nor for physical effort ($\chi^2 = 3.636$, $p = .16$). Table 6.8 contains the more detailed analysis results for both of the gestures. We observe there that *Clapping* scored significantly lower in condition Qx on intuitiveness when compared to conditions $Q1$ and Xp. All three conditions scored differently with respect to whether the subject would gesture in this way.

	condition	$Q1$	Qx	Xp		
	N	99	10	10	χ^2	p
Clapping	Intuitiveness	61.42	41.25	64.65	3.446	.18
	Physical effort	59.57	76.35	47.95	3.596	.17
	Would use it	60.31	40.20	76.75	5.878	.05
PinkieTrigger	Intuitiveness	55.23	78.30	88.90	12.157	< .01
	Physical effort	61.37	59.25	47.20	1.644	.44
	Would use it	55.15	78.00	90.00	12.619	< .01

Table 6.8: Differences between the online ($Q1$) and validation (Qx and Xp) conditions for the *context menu* gestures. Kruskal-Wallis H analyses results with mean ranks are reported. Insignificant results have been shaded.

The results from condition Qx show significant differences between gestures for the three questions. However, the results from condition Xp show no such significant differences apart from whether the participant would use this gesture. Similar results follow from conditions Qx and Xp although the results from condition Xp are not as significant. When comparing these findings to those in condition $Q1$, we see that *PinkieTrigger* scored higher on intuitiveness in conditions Qx and Xp but that the physical effort that is required and whether the participants would use it do not differ between these conditions. In both conditions *PinkieTrigger* was ranked as the best gesture for opening a context menu. However, the difference with *Clapping* was minimal.

For *Clapping*, our subjects found it hard to keep pointing accurately while clapping to open the menu. *PinkieTrigger* proved strenuous on the subjects' hands but our participants liked the way it mimicked the right-button on a mouse. Some users

proposed to pinch the tips of the thumb and the pinkie finger together to relieve some of the stress due to the required hand shape. Numerous alternative gestures were proposed. One subject proposed the use of other hand to perform actions that resemble right-button events. Another subject expanded that idea by suggesting the use of his preferred selection gesture, *ThumbTrigger*, with the non-preferred hand. It was also proposed to point with the whole hand while tapping the tips of the thumb and index finger for selecting and tapping thumb and middle finger for opening a context menu. The argument there was that those three fingers can be bent more comfortably. Another possibility is to rotate the wrist, either pronation or supination, to trigger the menu. It was also proposed to pinch the middle finger on two different places to mimic the left- and right-mouse buttons.

6.3 Summary of findings

In Chapter 5 we described an online questionnaire in which 26 gestures for issuing a total of six commands were evaluated by a large user group. The results from that condition ($Q1$) were validated in this chapter with two conditions of smaller scale. Validation was necessary because subjects who filled out the online questionnaire could not experience and appreciate the gesture-based interface that was proposed via videoclips. We wondered whether the subjects in condition $Q1$ could imagine what it really would be like to interact using the proposed gestures. The two validation conditions Qx and Xp were identical except for the users that took part in them. The validation conditions used a working prototype interface in which the gestures from condition $Q1$ were performed before assessing them. In condition Qx we randomly selected users that first filled out the online questionnaire. Our aim was to find out whether the ratings between conditions $Q1$ and Qx are consistent. The third investigated condition Xp encompassed subjects who had not filled out the online questionnaire. Our aim there was to explore the consistency of the ratings between users who had seen the gesture videoclips before and those who had not. Apart from these differences, all three user groups had a similar background.

There were some differences between the gesture ratings in the three conditions. However, we will show below that these differences are minimal and that they can easily be explained and rationalized. Some preferences were less pronounced in conditions Qx and Xp than they were in the online questionnaire. In most of those cases the preference did exist but it was not significant due to the limited number of subjects that took part in conditions Qx and Xp.

For pointing, we found no significant differences: in all three conditions the *Ray-casting* gesture was preferred to point at locations on the screen. However, the subjects from condition Qx did find *Ray-casting* more fatiguing than the participants in condition $Q1$ and Xp: having to keep one's arm outstretched for pointing is fatiguing for prolonged interactions. Fine tuning the act of pointing was demonstrated by using *Tap once* after initial coarse pointing with *Ray-casting*.

There was a difference between condition $Q1$ and conditions Qx and Xp concerning the selecting gestures. *AirTap* was liked overwhelmingly in the online questionnaire but in both validation conditions *ThumbTrigger* scored similarly. Our subjects liked the physical feedback that *ThumbTrigger* offered upon 'clicking' although they did not miss this form of feedback in *AirTap*. Although we found that *DropIt* and *Select other* were both preferred for deselecting in condition $Q1$, conditions Qx and Xp showed a significant preference for only *Select other*. Especially the fact that this gesture is familiar from existing WIMP interfaces led to this choice, hinting that our subjects prefer predictable and recognizable interactions. In addition, the users in conditions Qx and Xp commented that when using *DropIt*, it requires them to first make a fist before they can perform the deselect-gesture *DropIt*. This requires an additional step that broadens the gap in the 'gulf of execution' [152].

To resize objects condition $Q1$ showed little difference between *Fingers apart* and *Hands apart*. Subjects in conditions Qx and Xp preferred the latter, significantly. The difference in physical effort involved between the three conditions did not differ. Our subjects mentioned that when they had to resize more, the amount that the hands could indicate enabled much more precise resizing.

The most important difference between the three conditions with respect to the activate and deactivate gestures was that *ThumbTrigger*, like with selecting, scored higher on all accounts in Qx and Xp than it did in $Q1$. *Dwelling* was liked less in conditions Qx and Xp than it was in condition $Q1$, mainly due to what our participants called 'action by inaction': issuing an activation-command while holding the hand still felt inappropriate somehow. Although it has been shown that for hand-held devices dwelling is a suitable means of pointing [9], the lack of feedback during the dwell-time and the inactivity while pointing can lead to confusing interactions with the user: when will the system respond and why is it responding when I do not do anything? With respect to opening an options menu, we found similar ratings in all three conditions except for the intuitiveness concerning *PinkieTrigger*: significantly higher ratings were found in conditions Qx and Xp than in $Q1$. Subjects commented that it was difficult to determine when to clap and how to combine it with pointing.

6.4 Conclusions

In general our subjects found it better to use just one hand for gesturing because that was already fatiguing for prolonged interaction sessions. However, two hands offer an explicit means to indicate distances. Resizing is a prime example but for pointing with *Repetitive taps* it was proposed to indicate the start of the movement with one hand and to use the other to stop. We found that subjects found it hard to imagine why we included the activate and deactivate task: it was unclear what this task was supposed to do in existing interfaces they are familiar with. We consider it important that all subjects felt that they were in actual control during both condition Qx and Xp. This ensures that our findings are based on experience with a working

interface. A main contributor was the fact that pointing through *Ray-casting* actually worked and that the gesture could be detected robustly by the operator. Although some participants mentioned that they did hear the operator pressing buttons during the investigation, they did not feel hindered or influenced by it. We conclude that the results from both conditions Qx and Xp are comparable to those found in condition $Q1$. This entails that the findings from the online questionnaire, which are based on a large user group, can be used to define a gesture set for issuing commands to a large display with just the hands. In addition, this means that an online investigation can be used to get representative feedback for, in our case, finding out which gestures are suited for explicit command-giving to a large display from beyond arm's length.

6.5 Discussion

The prototype implementation that we used in the validation conditions was meant to offer our participants a feeling of what it would be like to issue commands through various gestures. We do not have an explanation why the subject preferences differed between conditions Xp and Qx, for example, in fatigue while ray-casting: we expected these conditions to score similarly due to experiencing the gestures while, in contrast, subjects in condition $Q1$ had to imagine the interaction. It is not expected that increasing the number of participants makes a significant influence. Subjects in Qx scored higher for their knowledge of smartphones and the iPhone but that too is something we do not expect influences these findings. It would seem that there are other factors, that we have no information about, that influence the preference for gestures in each of the conditions.

Although all participants felt in actual control of the interface, we observed that the glove with the IR LED for pointing did not always operate as intended. Although the set-up was calibrated for each participant, some users tried to point mostly by turning their index finger instead of keeping their hand perpendicular to the display so that the IR LED could be adequately observed. Although none of the users felt restricted by this shortcoming of the prototype, it might have influenced their experience. To increase pointing accuracy, jitter reduction is a requirement. König *et al.* [113] showed that an infrared laser pointer can be successfully employed to create a virtual cursor that can be stabilized using prediction filtering algorithms. An operator was switching between the application-states in our prototype. His responses might have also influenced the user experience when he did not time state-switches precisely or correctly. Although the number of interpretation-errors was not registered, our users did not mention to be influenced by those mishaps.

The execution of intended gestures by the subjects varied to some extent. Although all users were shown the intended gesture by both the videoclips and by the operator, in most cases they quickly started to improve the gestures for better comfort and, in most cases, minimal movements. We repeatedly mentioned to assess the gestures based on the intended gesture and not based on their variations. However, it might be that the findings in validation conditions have been influenced

by subjects not comparing the intended gestures but rather scoring based on their own improvements on the evaluated gestures. The reason for trying to minimize the gesture movement might be explained by increased comfort but also by the speed of gesturing. The keystroke level model, described by Card *et al.* [21], includes the time for homing the finger over a button to press. For *ThumbTrigger*, as an example, this can describe why small variations on the spot where the user tapped with the thumb were observed. Perhaps these users were trying to gesture as quickly as possible, even though we did not instruct them to do so. It might be possible that in doing so, the users tried to minimize the homing time.

Our findings in the three conditions have shown that *AirTap* and *ThumbTrigger* do not differ much as far as the user is concerned. Both gestures resemble the act of pressing a button, only the form in which the button is pressed differs. The physical feedback that *ThumbTrigger* provides combined with the minor improvement to pinch the tips of the thumb and index or middle finger together for clicking provide an even more pleasant, less strenuous and less invasive way to issue commands through gesturing. We expect that *AirTap* scored relatively higher in the online questionnaire (condition *Q*1) because there our participants did not experience what it is like to gesture in this way. The appeal of pressing a button is undeniably strong [76], mainly because buttons are pressed on a daily basis, even insofar that people press a button to see what it does: “oeh, what does this button do?” [36; 152].

Another contributor to a successful gesture-based interface is an easy and fast way to switch between interacting and resting. For example, our users mentioned for one of the deselect gestures, *Retract to rest*, that it took too much time to restart gesturing. Similarly, Baudel and Beaudouin-Lafon [5] showed in their early Charade system that fatigue, the non-self revealing nature of gesture-based interfaces and the lack of comfort of such systems inhibit the development of robust and widespread gesture interfaces. By facilitating an easy way to rest the arms, prolonged interactions might also be a possibility. One possibility is to introduce a tilted sketching table on which the user can rest his hands. The rest-stand might also contain some interaction functionality in the form of buttons [8], an manipulatable overview [32] and explanations on the workings of the system [215]. Another solution to facilitate rest to the arms while gesturing is to introduce ‘de-stressing movements’ that deviate from, for example, a constant pointing posture [146].

Chapter 7

Gestures in the Interface

> "The major difference between a thing that might go wrong and a thing that cannot possibly go wrong is that when a thing that cannot possibly go wrong goes wrong it usually turns out to be impossible to get at or repair."
>
> **Douglas Adams**
> British writer, 1952–2001 – Mostly Harmless, Picador, 2002, pp.114–115

A gesture set can be defined for explicitly issuing commands to an interface that observes and transparently reacts to a user's gesturing. The experiments that are described in the previous chapters have shown that users prefer to explicitly start and end their gesturing when they are giving commands to an interface. For example, the *ThumbTrigger* gesture, by Grossman *et al.* [64], allows the user to start gesturing by pressing his thumb on the base of his middle finger and end it by lifting the thumb again. Positioning the thumb over the correct finger for tapping it can be described with the keystroke level model and is known as 'homing' [21]. While gesturing, we found that by mapping the hand's movements directly to the interface, the interaction becomes transparent to the user. Elaborate hand shapes and movements can similarly be matched to an interface but recognition can then occur only after the gesture is (partially) completed [161; 171].

The main thing that was lacking from the previous experiments is an interface that is fully controlled by the user. The last experiment in Chapter 6 did enable the user to control the interface partially but there we observed that, at times, the interaction, which was partially induced by the operator, did not match the user's intentions. In this chapter we aim to correct this shortcoming by evaluating an interface that the user can fully control. As a starting point we use the gestures we evaluated in our previous chapters, among others, *ThumbTrigger* and *Hands apart*. The work in this thesis does not include building a system for unobtrusive gesture recognition, see also Section 1.4. We rather employ existing technologies and techniques for looking at users gesturing as much as possible in our evaluation of a gesture interface for command-giving to large displays beyond arm's length.

This chapter is structured as follows. We describe the method of our evaluation in Section 7.1: the tasks that we asked our subjects to perform, the commands that

they could issue, the devices that they used to interact with the large display and the software implementation of the prototype that we built. Results of our experiment are reported in Section 7.2: we report on data collected through observations, questionnaires and informal interviews. Our findings are summarized in Section 7.3. Conclusions and a discussion end this chapter in Sections 7.4 and 7.5, respectively.

7.1 Method

The user is interacting at a distance to the screen in this experiment. Gill and Borchers [60] describe three zones in front of the display that influence interaction: the action, negotiation and reflection zones. The action zone requires direct contact, touch, with the interactive surface. A user in the reflection zone is passive towards the display with no intent to act. We focus on the negotiation zone where the user is engaged with the display insofar that there are actions or indications of actions.

For this experiment we asked the participants to perform selected gestures repeatedly. In this way, they could experience what it will be like to interact with a gesture-based interface for a limited period, roughly 20 minutes, and imagine what it would be like to do so for a prolonged period. This gives us qualitative insight into gesture-based interaction and, more precisely, the users' perception of employing gestures in this interaction. In this section we will first describe the tasks that participants in this experiment completed, see Section 7.1.1. Second, we describe in Section 7.1.2 the time-schedule that we used to carry out the experiment. Third, Section 7.1.3 describes the gestures/commands that were available to complete these tasks. Fourth, the implementation details of the hardware that underpin the evaluated interactions are described in Section 7.1.4. Fifth, we describe the software details in Section 7.1.5. The first and third sections are plagued to some degree by the paradox of the chicken and the egg because we design the GUI based on the gesture-commands: the available commands form and limit the tasks that can be completed and vice versa.

We use questionnaires to evaluate the user experience while operating the interface. One questionnaire was filled out before starting each trial, see Appendix B.1, two were filled out after, see Appendices B.2 and B.3. The first questionnaire contains questions dealing with the experience that our participants already had with with gesture interfaces. The other two contain questions that address the overall interaction and the interaction for each command, see Section 7.1.3, respectively. Just as we did in the previous chapters, we first checked whether our data followed a normal distribution with a D'Agostino-Pearson K^2 analysis [33]. If the distribution was not normal, we looked at the cause for it in the kurtosis or skewness of our data that can indicate a ceiling or floor effect. Section 7.2 will show that a normal distribution was not found for all questions. We evaluated the ratings for the commands with a Kruskal-Wallis H analysis to find out if there were significant differences between commands. We then used a pair-wise independent samples Mann-Whitney U analysis to investigate what these differences are.

7.1.1 Semantics

We asked our participants to perform four randomized pattern-matching tasks on a large display. Each task consisted of finding a goal-state that was a certain orientation and zoom-level of a 3D mesh. We provided an image of the desired goal-state to the participant. This image could be referred to at any moment. As a starting point we offered the participant four different 3D meshes to choose from, see Figure 7.1. These meshes are biochemical structures that are used by, for example, biochemists, to discover function from the form of the structure. The participants were not required to have any knowledge of the biochemical structures nor of its visualization standards. In this way, we reduced the task to a more simple pattern-matching task [211]. The setup that was used in this experiment, and the graphical user interface that is part of it, will now be described. We describe the possible interactions in Section 7.1.3.

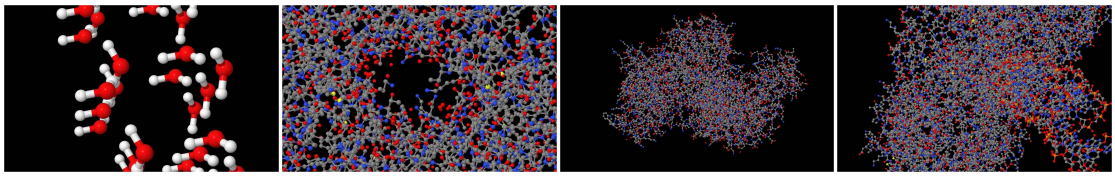


Figure 7.1: The four structures[1] that participants could open, participants did not require any knowledge on the structure.

Setup

The large display, sized 400 × 125 cm, used in this experiment is depicted in Figure 7.2. The display was created with two projectors that displayed a total resolution of 3840 × 1200 pixels; 1920 × 1200 pixels per projector. The projectors were mounted on the ceiling in such a way that the user did not cast shadows on the projection screen. Both projectors were set to the same color tone and they were calibrated to generate a single uninterrupted projection. The user was allowed to walk in front of the screen but was not allowed to come closer than 1.5 meters to the screen. This limitation, enforced with a line on the floor, meant that the user could not be at arm's length, or closer, of the screen. The way in which users interact with a display is influenced by the interaction zone in which they are located: action, negotiation and reflection [60; 169]. In this setup, the users were interacting in the negotiation zone.

[1]Four chemical structures are presented to the participants. Ice is a crystalline phase of water molecules, see Figure 7.1a. The 1a3n structure is better known as human hemoglobin and it can transport oxygen in humans [201]: we selected the structure without oxygen, see Figure 7.1b. The 1u04 is the structure of an Argonaute protein from Pyrococcus furiosus and it is thought to be implicated in mRNA cleavage during cell division [193], see Figure 7.1c. The 2f8s structure is an Argonaute protein from Aquifex aeolicus [240], see Figure 7.1d.

An infrared camera was used to look at the user's pointing behavior. The camera can detect the upper range of the visual spectrum, starting from 600 nm. We darkened the room in which the experiment took place to remove the presence of sunlight that might hinder the computer-vision recognition process [6; 93], see also Section 7.1.4. The room was uniformly lit with fluorescent lights. The examiner was sitting at the back of the room. He observed the user in his interactions and he made sure that the questionnaires were filled out completely.

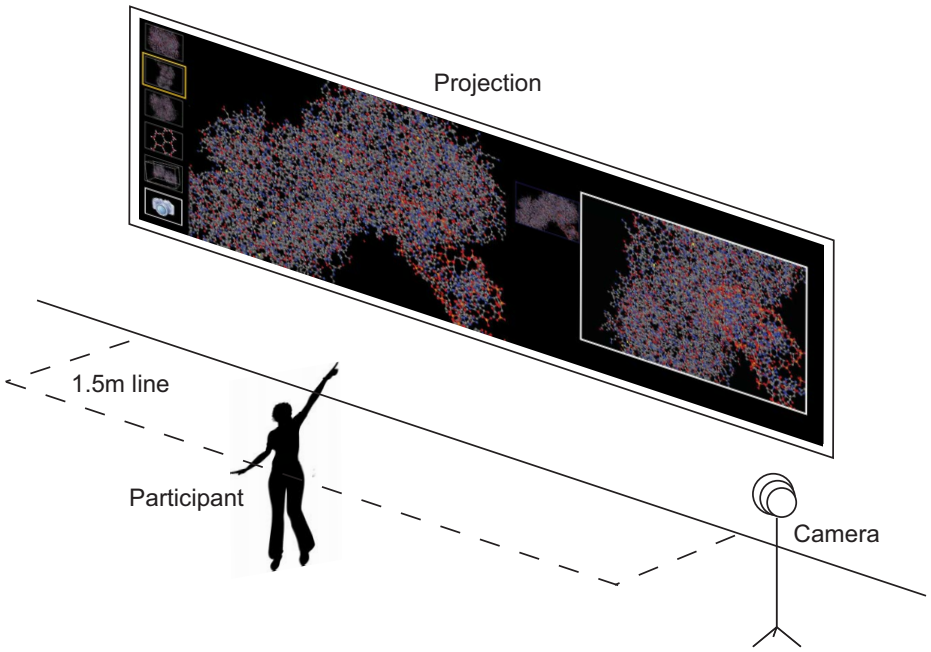


Figure 7.2: The setup used in this prototype. The camera stood at the back of the room, to the right of the participant; it could see the whole display. The participant could walk around but not stand closer than 1.5 meters to the display.

Graphical user interface

The graphical user interface that we used in this experiment is depicted in Figure 7.3. It consists of three borderless panels which are, from left to right, a context menu, a 3D mesh and a collection of 2D screenshots. Although we have called it a 'context menu' in the earlier chapters, in this prototype the menu is not context-dependent: it is just a menu from which the participant can select options. This menu contains six options. The first four options load a specific structure. The fifth toggles a bounding box around the structure so that its orientation becomes more apparent. The sixth option creates a screenshot of the current visualization in the central 3D panel on the right-most screenshot panel. The menu was visible continuously.

By selecting a biochemical structure from the menu, the 3D representation of that structure would be loaded in the middle panel with a set starting orientation and zoom-level. These settings were identical for all four structures. The structure could be rotated and zoomed in and out. We used Jmol to visualize the structures

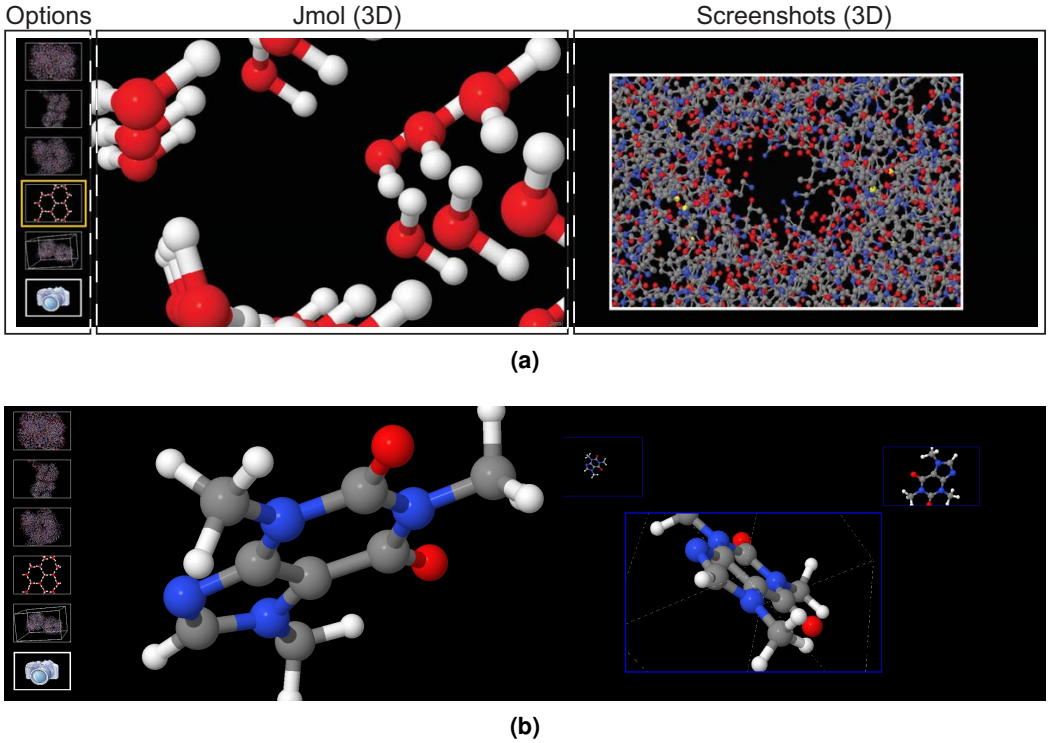


Figure 7.3: The Graphical User Interface used in this experiment, (a) an overview of the three panels in the GUI and (b) the start screen at the beginning of the trials.

in 3D. Jmol is an open-source viewer[2] for chemical structures in 3D. Jmol is mostly used for teaching and research in chemistry and biochemistry. Another 3D mesh would also have sufficed for the purposes of this experiment, for example, simple Tetris-blocks. However, we chose to use these biochemical structures because they are complex enough so that our participants would first spend some time searching for the correct structure and then some more time searching for the correct orientation and scale.

To enable the user to switch easily between previously visited locations we facilitated the use of screenshots that could be ordered as the participant saw fit. To ensure that all available commands were indeed repeatedly given by each participant, we requested the creation of at least two and deletion of at least one screenshot per goal that was offered. These screenshots were presented in the right-most panel. Each screenshot could also be loaded so that the 3D mesh that was its origin was again displayed. When a screenshot was created it would be sized roughly 10% of the total height of the display (125 pixels). However, to help recall the details of each screenshot, it could be resized as the participant saw fit. Screenshots could also be removed. The goal-state was represented as a screenshot that could neither be loaded nor removed.

[2] `http://jmol.sourceforge.net/`, 4 November 2009

7.1.2 Time schedule

We intentionally included multiple, similar looking structures so that the participant would have some trouble, firstly, finding the correct 3D mesh of the goal and, secondly, rotating and zooming it to match the required goal state. Our reasoning behind this was that the tasks are not time-crucial: we wanted the participants to perform all commands available to them repeatedly. The experiment required roughly 45 minutes to complete fully. We did not time each part of the experiment but we estimate that the practice session to get used to the interface took about 5 minutes while completing the tasks took a total of 20 minutes.

7.1.3 Commands

The best-scoring gestures from our previous experiments are the basis for this experiment. Each of the following commands was evaluated with a questionnaire in which we asked how easy it is to learn and remember the gesture ('1: easy to learn' - '7: difficult to learn'), comfort for the hands while gesturing ('1: cramped' - '7: comfortable') and we asked for additional comments, see Appendix B.3. In addition to these detailed questions, we asked the participants questions on the design of the devices that they used, see appendix B.2. We also closely observed the interactions.

Out-of-range and tracking

Ray-casting is used to detect whether a participant was in the out-of-range or in the tracking state. When participants pointed at the display with one or both of their hands, they were in the tracking state for that/those hand(s). Each hand could be tracked separately. We distinguished the participant pointing at each of the three panels with one or two hands. It was possible to simultaneously point at two different panels.

Select and deselect

We had to omit *AirTap* from our evaluation due to the technical inability to robustly detect this gesture, even though it scored as well as *ThumbTrigger* in the evaluations reported in Chapters 5 and 6. Instead, we included both *ThumbTrigger* and *Pinch* in our evaluation. *Pinch* is a variation on *ThumbTrigger* that was not present in our previous evaluations. We chose to include *Pinch* in our design because it was found there that slight user-dependent variations existed in the execution of *ThumbTrigger* to increase comfort levels while gesturing. The device we used to detect gesturing, see Section 7.1.4, was fitted on the index and middle fingers at the beginning of each trial. We then allowed each participant to decide the most comfortable gesture in a brief practice session. The position on the hands where the participants tapped with their thumb was marked for further analysis, see Appendix B.2. Despite the expected small variations that each user may decide upon, we use *ThumbTrigger* in the remainder of this chapter as the name for the preferred gesture made by each participant to increase readability.

It was possible to select and deselect in the menu and screenshot panels; however, the meaning of this was different for each panel. In the menu panel, each of the six options could only be selected. In the screenshot panel, the user could select a screenshot to restore it to the 3D panel. Selecting screenshots or menu options could be undone by selecting another option or screenshot. However, if no screenshot was made of a particular 3D orientation and zoom level, the participant had to start anew. For deselecting, *Select other* was opted given the positive feedback that we had received on it. Note that deselect could only be performed on screenshots. The bounding box option could be toggled which meant that by performing *ThumbTrigger* on it a second time would deselect the bounding box.

Rotate

A special selection case is rotating in a 3D visualization. The participant performed *ThumbTrigger* with one hand on the 3D panel to rotate the biochemical structure to the desired orientation. It was possible to rotate around the x and y axes. It was possible to rotate around the viewing axis (the z axis) using *PinkieTrigger* with one hand. We added this gesture based on user comments in a couple of exploratory trials. This approach to rotating, which is known as ArcBall [83], is typical for the rotation schemes that are used in 3D design and drawing applications such as Autodesk AutoCAD® and Adobe Photoshop®. This action can be undone by rotating in the opposite direction(s). Arcball does not allow for rotating around the z axis. We allowed an ArcBall rotation around the z axis by performing *PinkieTrigger* while moving the laser dot horizontally.

Resizing: shrinking and enlarging

The prevailing gestures for the resize command were *Hands apart* and *Fingers apart*. Given the scale of the large display and the comments we noted in the previous evaluations, we chose to focus on *Hands apart*. However, like *AirTap*, there was no means to detect the start and end of this gesture robustly, see Section 7.1.4. Participants performed *ThumbTrigger* with both hands to signal the beginning and ending of their resize gesture. The users pressed both of their thumbs down for the duration of this gesture.

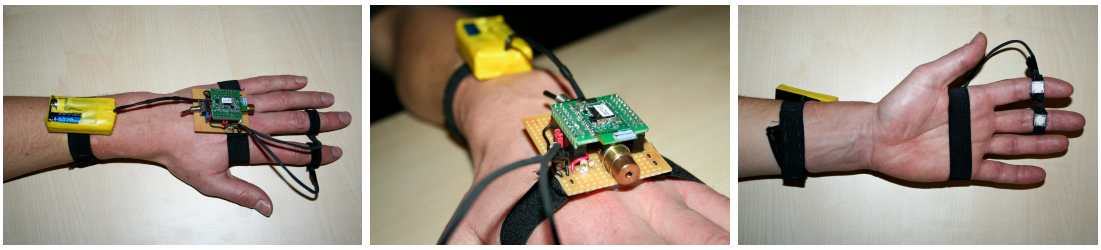
Resizing was possible on both the 3D panel and on the screenshot panel. For both panels, the participants moved both hands on the target, structure or screenshot, that they wished to resize. By performing *ThumbTrigger* and moving the hands apart for enlarging and towards each other for shrinking, the participant could resize the target to the desired size. Clearly, to undo a resize command, the participant needed to perform the opposite command, thus resizing the structure or screenshot to its previous size.

Restore and Remove

Restoring a screenshot could be done with *PinkieTrigger*. By performing this gesture, the structure with the orientation and size depicted in the screenshot would be restored in the 3D panel so that the participant could continue manipulating it from there. Lastly, screenshots could be removed by performing a *PinkieTrigger* gesture with both hands on them. The participant pointed at the target that they wished to be removed. After performing *PinkieTrigger*, the screenshot was removed from the screenshot panel. It was not possible to undo this action.

7.1.4 Devices

The state-of-the-art for detecting, recognizing, interpreting and, equally important, reacting to the participants' gesturing in an unobtrusive way, mostly using camera-based solutions, is far too immature for our purposes [168]. After all, the gestures that we evaluate in this experiment are fine-grained gestures in which minor changes in hand shape and bending of the fingers convey the gesture's meaning. In addition, such approaches are ill suited for the type of fast-prototyping evaluations that we wish to perform to evaluate the user experience in interactions by means of gesturing. We therefore designed and built a pair of wearable, wireless devices that allowed us to evaluate the gestural interactions described in Section 7.1.3.



(a) (b) (c)

Figure 7.4: The glove device: we used two of these. The view (a) of the top of the hand, (b) of the front of the hand and (c) of the hand palm, note the buttons on the fingers.

Figure 7.4 depicts the devices that we used in this experiment. In the remainder of this chapter we will refer to these devices as gloves. The first thing that catches the eye is the laser that is attached to each glove. We used these lasers for pointing towards the display. Myers *et al.* [141] investigated how users could comfortably hold a laser pointer in their hand but here we attached the laser pointer to the back of the hand so that the whole hand can be used for pointing [28]. In this prototype we used red lasers (650 nm, 1mW, laser class 2) so that the user was directly aware of where he was pointing. König *et al.* [113] used infrared lasers for this task so that the users could not see where they were actually pointing. That fact is used to display a purely digital 'laser' dot that can be stabilized by means of jitter reduction. At the time of the construction of this prototype we did not have such infrared lasers available. Both hands had a laser with which the participant could point at

the display. We calibrated the location of the display by marking its corners in a separate calibration-phase that was performed before the participants were invited to start their trial.

It is apparent from Figure 7.4c that each glove is equipped with two buttons that the participant can use to perform a *ThumbTrigger* gesture. The buttons are sewn on elastic rings that could be placed on any finger at any position or orientation on that finger. By allowing the positioning of these rings we wished to evaluate the most comfortable spots on the left and right hands where the user would use *ThumbTrigger* or *Pinch*. Another reason for allowing the user to place the buttons as he saw fit is found in the keystroke level model [21]. The time it takes to perform a *ThumbTrigger* or *PinkieTrigger* is partially described by the time needed for homing the thumb to the correct starting place for that gesture. We wondered in the previous experiment if slight nuances in the *ThumbTrigger* gesture might be caused by this model. By allowing the user to place the buttons as he saw fit, we tried to prevent the user from performing as quickly as possible; he performed rather at his own pace. We could detect a user pressing, holding and releasing each button separately and for both gloves simultaneously. Note that the A button corresponds to the *ThumbTrigger* gesture while the B button corresponds to *PinkieTrigger*.

The buttons are connected to a chip with a Bluetooth connection so that we could read button presses wirelessly. This was a design requirement because in preliminary tests with a magnetic flock of birds sensor we noticed that the users felt hindered in their interaction by the wires that connected them to the interface. In addition, this made them more aware of the limitations of the interaction which, in turn, restricted their immersion in the interface [219]. Both the laser and the Bluetooth-enabled chip were strapped to an elastic band that the participants wore around the palms of their hands. Powering the lasers and Bluetooth chips was done with two alkaline batteries sewn to an elastic wristband, see Figure 7.4.

7.1.5 Software

Our software implementation consists of rather elaborate collection of components, see Figure 7.5. It can be broken down into three main components. First, the computer-vision analysis of the laser dots. Second, the button presses on both gloves. Third, the graphical user interface itself that has already been described in Section 7.1.1. Figure 7.5 depicts these components.

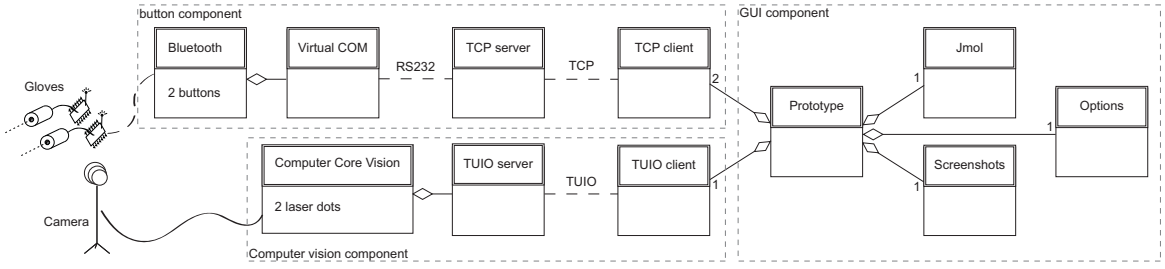


Figure 7.5: The software components of our prototype. Dashed lines separate components that run autonomously and on different computer systems.

The lasers were detected using the Computer Core Vision (CCV) open-source package that, with minor tweaks to the camera settings, allowed us to detect and track multiple laser dots. We marked the corners of the large display explicitly in a calibration process [53]. This allowed us to translate the coordinate system of the camera images (352×248 pixels) to that of our Jmol interface ([0.0 - 1.0] for width and height of the display). The location of these laser dots was translated to the TUIO protocol for multi-touch interactions [97]. Although we did not use this protocol for what its design intended—multi-touch—we did find it to be an adequate means to communicate the laser dot locations to the rest of the application.

Both gloves communicated button presses through the Bluetooth protocol. Repeatedly starting and stopping the Bluetooth connection strained the programming on the Aircable chip so that the connection became instable. Therefore, we created a virtual serial port on the computer so that the Bluetooth packages could be read when received. We used a TCP socket connection to establish and maintain the Bluetooth that our interface application could listen to and process the button presses on both gloves.

The prototype interface continuously received signals of buttons being pressed and released on the gloves in addition to the laser dot positions. However, it only responded when a button was pressed or released. Depending on the location of the laser dots, and the button(s) that was/were pressed, the interface would, for example, start resizing a screenshot or it would toggle the bounding box.

7.2 Results

Here we report the results of our experiment with this prototype. First, our sample is described in Section 7.2.1. Second, Section 7.2.2 describes the experiences that we obtained, both through our questionnaire and through observation.

7.2.1 Sample

A total of twenty-three subjects participated in this within-subjects design. All participants studied at or worked for our university. Participants were 29 years old on average (ranging 24-47 years, $\sigma = 5$ years). All participants completed the experiment. Five participants were female, eighteen were male. Eight subjects held a Bachelor's degree, thirteen subjects held a Master's degree and two a PhD degree. All participants were right-handed. Two participants had taken part in our Wizard of Oz experiment (Chapter 4), eighteen in the online condition ($Q1$) and nine in the two validation conditions (Qx and Xp) of our large-scale experiment (Chapters 5 and Chapter 6). One subject was familiar with the structure of ice, but had not seen the other three structures before taking part. All other participants were unfamiliar with the four structures that were used in the prototype.

Figure 7.6 shows the results of the ratings of our participants' knowledge related to gesture interfaces. Participants were moderately familiar with pen-based devices

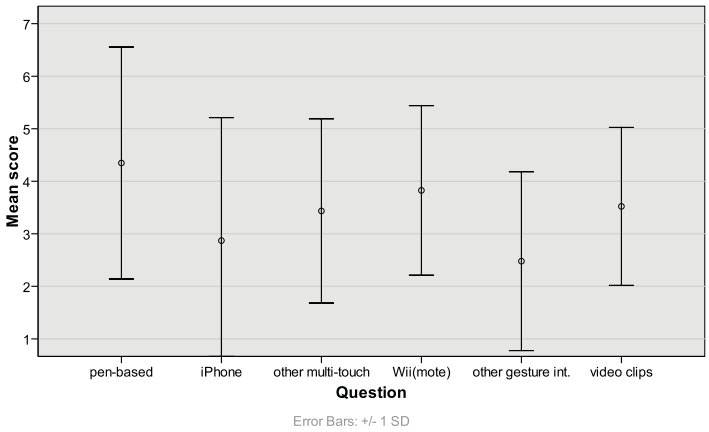


Figure 7.6: Experience of our subjects before taking part in the experiment.

such as a PDA and tablet PC and they also mentioned the Nintendo DS, cellphones and the Apple iPhone in this category. The participants were not familiar with the Apple iPhone but more so with other multi-touch systems. They mentioned the touch tables at our research group [44], the Apple iPhone itself, the Apple Touchpad and the trackpad on their notebook. Our participants were moderately familiar with the Nintendo Wii and its Wiimote controllers but less so with other gesture interfaces for which they mentioned the Playstation EyeToy, data gloves, photoplay, the Personal Space Station [140] and Firefox mouse gestures. Our participants were not so familiar with video clips of gesture interfaces. 'Minority Report' was mentioned explicitly nine times while other sources were 'The Island' (2), 'Paycheck', 'Star Trek' (2), 'Iron Man' but also Oblong's G-Stalt[3] and Microsoft's Surface multitouch table. Other gesture interfaces that were named included: 'camera-based interfaces', 'gesture detection in large rooms such as waving and pointing', 'endoscopic operation robot in surgery', 'EMG-based guitars', 'Microsoft Natal' and, again, 'Firefox mouse gestures'. A D'Agostino-Pearson K^2 analysis showed that there are normal distributions for these ratings except for experience with other gesture interfaces ($K^2 = 9.860$, $p < .01$) than Nintendo's Wii, see Table B.1. This deformation is a result of a high values for skewness and for kurtosis.

7.2.2 Experiences during the experiment

Here we describe the experiences of our participants during the experiment. First, we describe the results that we obtained from our questionnaires. Second, we describe our observations during the twenty-three trials.

Questionnaire overall

A D'Agostino-Pearson K^2 analysis showed that the ratings for the whole interaction do not follow a normal distribution, see Table B.2. Figure 7.7 depicts the results.

[3] `http://oblong.com`, June 16th, 2009.

We can see that the overall experience was positive. Our participants understood how the lasers were used for pointing, the pointing accuracy, operation speed and comfort while interacting were high, and there was limited fatigue in the hands and arms while interacting. The rating for the 'fun-factor' was high as well. The smoothness of the interaction scored somewhat lower. There was only one participant who explicitly commented that the interaction could have been smoother. Three participants mentioned that 'Getting used to [it] is difficult because the lasers have the same color' by which they meant that they at times had difficulties in determining which laser dot originated from where. On that respect, it was also mentioned that 'Inaccuracy was not so much a bother because you get visual feedback from the interface *and* the lasers'.

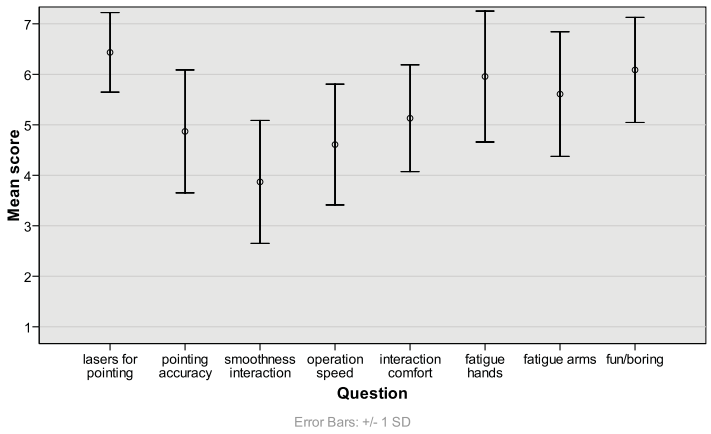


Figure 7.7: Overall interaction ratings.

Figure 7.8 shows where the buttons were placed on the participants' hands. Although we placed the buttons in the middle of the index and middle fingers at the beginning of each trial, after the practice session, we asked each participant if the buttons were placed comfortably and, if not, how they would prefer to place them. Only three participants decided to change the buttons when so asked. Others did so of their own accord, mostly because the rings were either too wide or too narrow. This was especially true for the five female participants, due to their slender fingers: they slid the rings down as far as needed to keep them from falling off entirely.

The buttons could not always be placed at the place on the fingers where the users first intended to place them. Because of their slender fingers, this mostly occurred with our female participants. The rings on which the buttons were sewn are made of elastic band. The rings were designed to not be too tight but the difference between the thickness of the fingers was not anticipated. Our participants did not mention that this caused the gesture to be uncomfortable. We found no significant difference for the comfort between men and women. Our female participants did rate the perceived operation speed significantly higher as the male participants did ($p = .02$), see Figure 7.9.

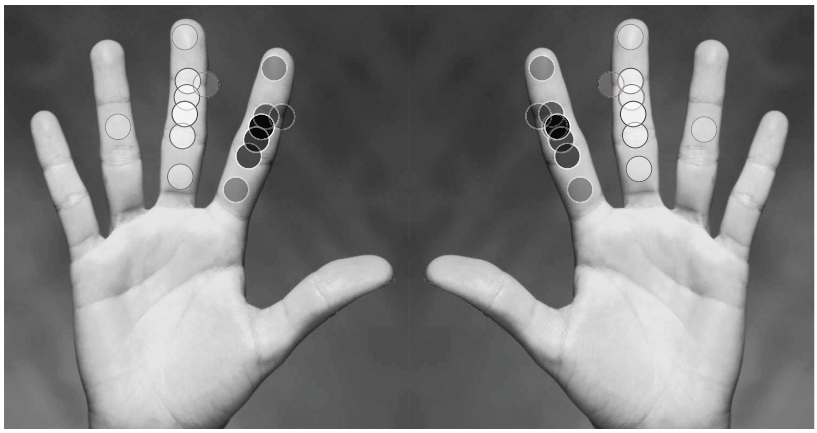


Figure 7.8: Button placement on the hands. The black dots represent the A button while the white dots represent the B buttons. The more intense the dot, the more participants placed a button there.

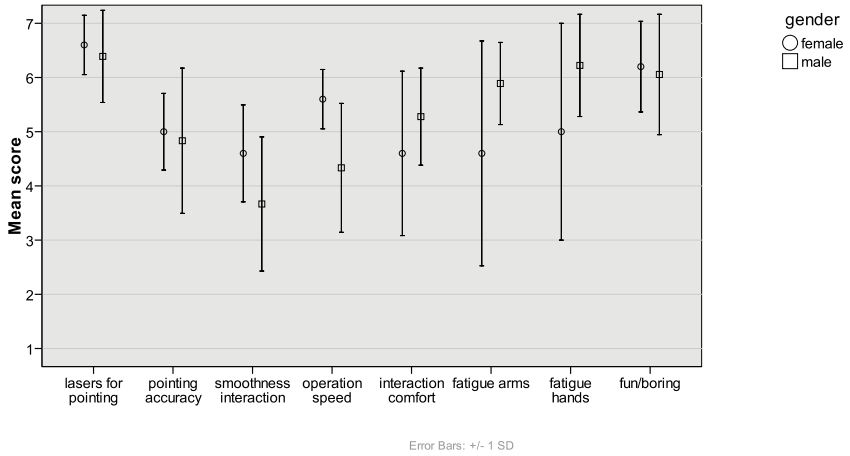


Figure 7.9: Overall interaction ratings per gender.

Questionnaire per command

Figure 7.10 depicts the ratings for each of the seven commands that we evaluated with a questionnaire. This figure shows that the ratings to learn and remember a gesture and for the comfort in gesturing scored similarly for all seven commands. A D'Agostino-Pearson K^2 analysis shows that the ratings per command do not follow a normal distribution: for learning and remembering a gesture we found $K^2 = 41.9$ ($p < .01$) and for the gesture comfort we found $K^2 = 42.5$ ($p < .01$). This deformation is caused by high values for kurtosis (1.7 and 1.4 respectively) and high negative values for skewness (-1.4 and -1.3 respectively). Table B.3 shows the ratings for each command separately. A Kruskal-Wallis H analysis shows that there is a significant difference between ratings for the seven commands with respect to how easy they were to learn and remember ($\chi^2 = 36.466$, $p < .01$) but not for the comfort of performing the gesture ($\chi^2 = 8.125$, $p = .23$). We performed an independent samples analysis on the seven commands using a Mann-Whitney U analysis for the question how easy it was to learn and remember the gesture. Rotating and

resizing the structure (3D) both scored significantly higher than moving a screenshot (2D) and than selecting options. Moving and resizing a screenshot (2D) scored significantly lower than restoring a screenshot and than deleting a screenshot. Resizing, restoring and deleting a screenshot scored significantly higher than selecting options.

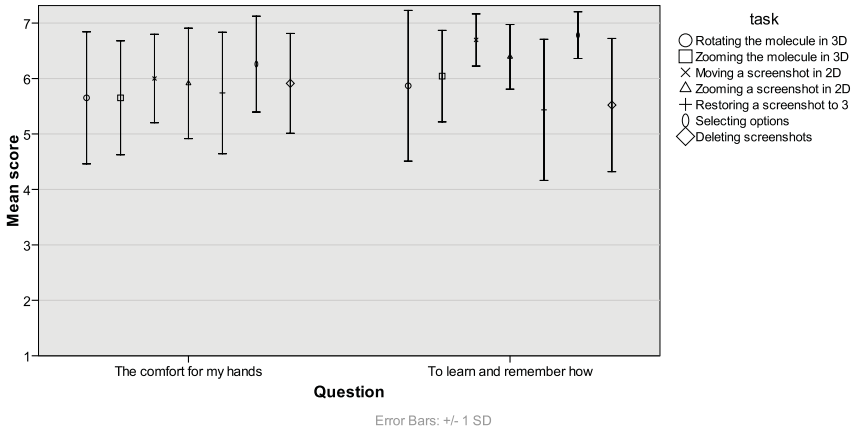


Figure 7.10: Detailed interaction ratings.

Two participants commented that, for rotating the structure in 3D, there was an irritant jitter when rotating which they attributed to low resolution of the pointing device. In addition, the reaction time of the devices was deemed 'too slow for my actions'. Another participant found this approach intuitive because it is based on the traditional mouse-based control of 3D space. For resizing the structure in 3D one participant wondered if the method for resizing was absolute or relative. He could not figure this out which was confusing for him. Three participants mentioned that small changes in pointing could lead to big changes in resizing the structure. One participant mentioned the Apple iPhone as the source of this gesture. Four participants mentioned that the response time was too high.

For moving screenshots in 2D, four participants mentioned that the calibration is very important and that it should be better calibrated because 'coupling [the laser dot to the screenshot] was clumsy'. Another participant mentioned that this method was very easy to understand: 'point and click, how much easier can it get?'. Resizing screenshots received the criticism that it should have worked the same way as in 3D (1 participant). Three participants found it hard to find the correct spot for both laser dots to start resizing. One participant mentioned that the response time was too high and another participant wondered which button to press.

Restoring a screenshot to 3D with *PinkieTrigger* was deemed unintuitive by one participant who preferred to just drag the screenshot to the 3D panel. Seven participants mentioned that they did not use this gesture much. For selecting options from the menu, one participant mentioned that they had accidentally selected options while performing resizing in 3D because she came too close to the menu with a laser dot. Another participant mentioned that additional feedback mechanisms, for example, an audible click, would be nice although the highlighted box in the options

menu also helped him considerately. With respect to the two-handed *PinkieTrigger* gesture for deleting, three participants mentioned they felt that they had not used it much. Another participant mentioned that it would be more fun to throw it away into a trash bin but that *PinkieTrigger* was easier. One participant mentioned that, for all comfort questions, he would have scored higher if our gloves would have been smaller 'like a [..] ring'.

Observations

The most prevalent posture for our participants to stand in was with their upper arms along their body and both their lower arms pointing towards the screen, even when they were only actively using one or even neither of the hands. When asked why they did not stretch their arms for pointing they commented that it was the most comfortable way for them to stand. It was rare for participants to walk around in front of the display although we did explicitly explain to them that it was allowed as long as they did not cross the 1.5 meter line. We did notice that all participants were switching the leg on which they were standing to stand more comfortably.

When performing the pattern matching tasks, most participants first loaded each of the four molecules to discover what they were looking at. After this exploration stage, they started manipulating the structure to fit the requested goal. We noticed that the subjects frequently mixed up the structures 1a3n, see Figure 7.1b, and 1u04, see Figure 7.1c. The target state for 1a3n (Figure 7.1b) was the hole in its center and the starting location of 1u04 seemed to have a hole in it.

Almost none of the participants noticed that they switched hands for pointing, between their left and right hand. When asked why they did so, they were at first surprised to find out that this was the case after which they mentioned that it was the most comfortable way for them to point. One participant commented that she was 'very right-handed' when performing the tasks although we observed that she too was switching her left and right hands for pointing. We did not observe any participant always using the left hand to point to the left side of the display, or vice versa. All participants mentioned, when asked, that they liked the visual feedback that the laser dots provided to them. They also argued that it was clear that when they did not press a button, the interface would not respond. One participant preferred to have the laser attached to his fingertip but the other participants frequently mentioned that they liked pointing with their whole hand: they argued that it was more comfortable to keep their fingers relaxed.

One participant had significant difficulties in perceiving depth in the Jmol panel while two other participants suggested that perception of the 3D structure could be improved by using 3D goggles [211]. One participant mentioned that she felt that the response time of the interface was high but that she accepted it because it was a new type of interface. Were this to happen on her PC, it would be totally unacceptable.

7.3 Summary

The gestures that were preferred in our earlier experiments, see Chapters 4, 5 and 6, were evaluated in this chapter. We built a working prototype with two wireless, glove-like devices that enabled our participants to interact beyond arm's length with a large display through gestures such as *ThumbTrigger*, *PinkieTrigger* and *Hands apart*. Our subjects experienced this interaction for twenty minutes after a practice session of five minutes. By giving our subjects a chance to interact for this amount of time we obtained qualitative feedback on the interactions. A set of four pattern matching tasks was given to our subjects. These tasks were designed in such a way that they required the subjects to repeatedly give commands to the interface to achieve the required goal. An image of a complex 3D mesh (a biochemical structure) was presented that the subjects had to reproduce by rotating and resizing a 3D mesh. In addition, the user could manipulate that image, and other images that could be made in the process of finding the requested target state, by moving and resizing them.

7.4 Conclusions

We found that all participants enjoyed giving commands through gesturing in our interface. They experienced gesturing as accurate, fast and comfortable. There was no fatigue in the hands and arms to speak of even though our participants tended to keep their arms tensed for the entire duration of the trial. The smoothness of the interaction could have been better which manifested itself mainly in rotating and zooming the 3D mesh. This was caused by the 3D rendering software that we used which was not fully customizable to our needs. Our participants preferred to shape the *ThumbTrigger* and *PinkieTrigger* gestures to fit their own comfort, placing the buttons that we used to detect the thumb pressing against another finger so that it was most comfortable for them. This mostly meant that the subject had to minimally bend his finger so that he could give a command with minimal effort. Women could not always place the buttons as they desired because the rings did not fit tightly enough on their more slender fingers. However, this was of no influence on our findings for comfort, accuracy, smoothness and fatigue. The combinations of gesture for giving a specific command were easy to learn and remember for the duration of our trials.

We can conclude that the gestures that we evaluated in our earlier experiments are fun (see Figure 7.11), comfortable and efficient for giving commands to a large display beyond arm's length. A wearable device was used in this experiment to robustly detect the gestures. None of our participants mentioned they felt uncomfortable to wear such a device even though it had to be tightly strapped to the subject's hands and arms. We suspect that a smaller device, which would still be attached to the back of the hand, might be more comfortable still. It has been argued in HCI literature [93; 168] that unobtrusive gesture recognition is a desirable way to interact through gesturing with an intelligent environment. However, we argue

Figure 7.11: One user out of 23 who is having fun during the experiment.

that by giving the user an explicit means to interact, for example, through buttons on a small wearable device, the interface will be more transparent for the user. In addition, holding or wearing a control device is an explicit signal to each user as to who is in control of the display [138]. This is, however, a topic for further research.

7.5 Discussion

Some participants struggled somewhat to distinguish between the two laser dots. It was not clear to them from which hand the laser dot originated. When giving a command this would on occasion give unintended and unexpected results. Apparently, our users mainly focused on the laser dot for feedback rather than on the spatial orientation of their hands pointing. König *et al.* [114] used an infrared laser pointer for pointing. By removing the laser dot entirely, however, the user is unaware of where he is pointing, let alone with which hand. König *et al.* therefore displayed a digital laser dot on which they could perform jitter reduction. This helped to stabilize the jitter effect caused by the hands trembling slightly in the interaction. Our participants commented that this effect was irritating at times. By replacing both our red lasers with infrared lasers, we can also represent each laser with unique color or shape. In the same way, multiple users can also be simultaneously supported without confusing as to who is doing what [99]. Note that in order for this to work, a robust tracking algorithm must be employed for detecting and tracking the laser dots on the screen, for example, by using a Kalman filter [224].

We evaluated how well our participants learned and remembered each gesture in only one trial per participant. However, it would be interesting to find out how well these gestures are remembered over the course of several days and weeks. Bieg [9] investigated earlier whether a better understanding can be gained of how easy it truly is to remember the gestures in the interface by intermitted learning by repeatedly revisiting the interface. A significant, albeit small, increase in performance

was found. However, the investigated task set, as defined by [91, Appendix A], requires sessions of at least two hours to complete. To prevent users getting 'gorilla arms', see Section 3.4.2, we would rather argue to investigate how well the gesture is remembered through qualitative analysis, such as performed in this experiment.

Part III

Conclusions

Chapter 8

Conclusions

> "Out with friends and last orders have been called in the pub. The alpha male of our group pulls out a stack of taxi numbers scrawled on old business cards. None of the firms is close enough. 'Richard has a new iPhone - let's try that,' my wife suggests. I pull up an app called AroundMe, which tells me where the nearest cab company is. Thirty seconds later and the taxi is on its way. My friends look on in envy and admiration. Alpha male looks despondent. 'I am part man, part computer', I tell myself."
>
> **Richard Fisher**
> New Scientist news editor – NewScientist issue 2272 "Appland: How smartphones are transforming our lives"

John Anderton, the lead character in the science fiction movie Minority Report, is used to interacting with wall-sized interactive surfaces by gesturing with his hands. The gestures that he uses to command the display look easy to understand, learn and remember. In Minority Report, we see John gesturing to point out, pick up and throw away items projected on the display. All this is happening because John is trying to solve a crime with this gesture interface. As far as the audience is concerned, John's interaction looks believable; mainly because the interface responds in a predictable way to each gesture-command. Gesture interfaces and, more specifically, the gestures that are used to give commands to them are what this thesis is all about. We studied gesture interactions at a distance, beyond arm's length, from the user: it is either not allowed or not possible to touch the display itself. Our experiments focused on gesturing with the hands, excluding speech as the prevalent additional input modality in multi-modal interfaces [119; 194]. We looked at these interfaces from a human point of view so that the interaction best suits the user's intentions of what she is trying to achieve with the interface [152].

Gesture interfaces that are operated at a distance can be applied in display-rich environments. We mention just some examples of these environments that are packed with displays that are on the walls, floor and embedded in the furniture. In smart meeting rooms, scientists analyze and interpret complex data structures such as they occur in life science research projects [178]. In shopping areas, display windows try to catch the eye of passers-by through interactive product information

[139; 215]. The surgeon's hands must remain sterile for the duration of the surgery in operating rooms of the future that facilitate easy access to a patient's information [216; 217].

The aim of this work is to explore, from a perspective of human behavior, which gestures are suited to control large display surfaces at a distance; why that is so; and, equally important, how such an interface can be made a reality. In an effort to understand gesture interactions with large displays at a distance, we performed four investigations (Chapters 4–7) that explore this type of interaction.

The rest of the conclusions chapter is structured as follows. First, we will report the findings of our four investigations in Section 8.1. Each reported investigation provides the basis for the one that follows. The understanding of which gestures are suited to be applied in gesture interfaces, and why, grows with the sequential progress in these investigations. In addition, gesture interfaces were built that allow us to formulate requirements for building gesture interfaces. Second, we reflect on the reasons for our findings with a discussion in Section 8.2 in which we introduce high-tech citizens as the main potential user group of the gesture interfaces that we built. This thesis is concluded in Section 8.3 with a brief outlook on the possibilities for continuing this line of research. We describe how speech might be added as an additional input modality and we introduce the design of a new device that is worn like a ring. This design builds upon the findings from this thesis. With it, we aim to investigate whether users prefer such a small wearable object to gesture with or whether they prefer to gesture without it; the latter will require a camera-based solution to detect gesturing.

8.1 Findings

Gesture interactions at a distance, as they are interpreted by the user, have been described in a four-state model before we started to study these interactions. The model describes when the user is out-of-range: when he is not interacting. It describes the state in an interaction where the hands are being tracked in addition to states where the user is selecting or manipulating the contents of the interface. State transitions occur when the user gives an explicit command: each command is given with a gesture. This model provides us with a basis to describe our gesture interactions. With this model, we have found in our investigations that these gestures are preferably used for issuing more than just one command, for example, using *AirTap* to both select and activate, depending on the current state of the interaction. Users like to reuse gestures in this way to keep things simple while interacting [237].

We have, as a first experiment, explored gesturing by users that received instructions on the command that they should give with a gesture while not receiving instructions on *how* to gesture. We did so to learn which gestures users make of their own accord and, equally important, why that is so. We found that gestures are started and stopped explicitly by changing the hand shape from rest to tensed and back, respectively. In addition, there was great uniformity between the gestures that different users made. The main difference between the gestures that we observed

was the tensed hand shape, in the gesture stroke phase [38], that was used while gesturing. We also learned from our users that a large influence on the choice for certain gestures was caused by their knowledge of technological developments and indoctrination. More explicitly, the Apple iPhone and the Windows-Icons-Menus-Pointing (WIMP) metaphor were mentioned repeatedly as such. Users felt as if they were in actual control of the display; immersed in the interaction that they found believable.

For our second experiment, we made an inventory of gestures for issuing commands to an interface from HCI literature, science fiction movies, existing (gesture) interfaces and everyday life. A command is given by making the gesture for it: we selected some gestures to issue more than one command. A large user group then evaluated, with an online questionnaire, the suitability of each gesture for issuing that command. Participants evaluated each gesture based on a videotaped example of the gesture with an interface that responded transparently. The preferred gestures are characterized by familiar actions that require as little effort as possible, for example, mimicking pressing a (mouse) button to select or moving the fingers apart to resize. The indoctrination by both traditional WIMP-style interfaces and recent mainstream multi-touch interfaces swayed our participants' preference towards these gestures.

The third investigation was a validation of our earlier findings because the users had previously merely imagined what it would be like to gesture as proposed. The gestures selected through the online questionnaire were validated with a prototype interface in which the user could point through ray-casting but where the start and end of a gesture were detected by an operator. Users with a similar background evaluated the suitability of each gesture to give a specific command by experiencing it. We found only minimal differences between the evaluations of the gesture from the online setting and after repeatedly experiencing it. These differences were mostly caused by users preferring to rest their hands as much as possible to increase the comfort while interacting. The users found it, above all, *fun* to interact through gestures in a seemingly working gesture interface. We also learned that participants consider it important that switching between resting and interacting is both easy and fast to do. The preferred gestures again show that there is a strong preference for gestures that mimic the pressing of a button. This evidence also supports evaluating gesture interactions through video prototypes as has been argued for by Tognazzini [203] and Bardram *et al.* [4]. Video prototypes offer a relatively fast means to evaluate the workings of a mature interface without having to build it.

To consolidate our previous findings we designed, built and evaluated a gesture interface with which the user can interact with 3D and 2D visualizations on a wall-sized display. This fourth experiment also provided us with experience in building and working with an operational gesture interface. Interactions consisted of the gestures for issuing commands that were preferred in our previous experiments. Again, we found that our participants preferred to interact with the least amount of effort and with the highest comfort possible. There was little variation between users in the shape of the gestures that they preferred: tapping the thumb on one of the other fingers, known as *ThumbTrigger* [64], was the prevalent gesture. Figure

8.1 depicts the prevalent gestures that we found in our studies: *Ray-casting* for pointing, *ThumbTrigger* was preferred for selecting objects and menu items, *Fingers apart* combined with *ThumbTrigger* for resizing, dragging with the thumb pressed down in *ThumbTrigger* for rotating in 3D and dragging in 2D, and *PinkieTrigger* for alternate selection commands.

Based on these findings, we conclude that it is possible to design and implement a set of gestures for giving commands explicitly to large display interfaces at a distance. These gestures are easy to learn and remember, they are comfortable to perform and they are fun to use. The design of this gesture set is based on actions that are familiar from everyday life as well as on actions that are found in both prevalent and emerging computer interfaces, for example, the Windows-Icons-Menus-Pointing metaphor and the iPhone. We argue that the main guideline in designing a gesture set for a gesture interface is that the gestures, as far as the user is concerned, should: be easy to learn and remember, be comfortable, require minimal movements and allow fast switching to and from a resting state. The aspects of this guideline also encompass the reuse of gestures in a different context, the familiarity of gestures from everyday life and the self-explaining nature of the gestures.

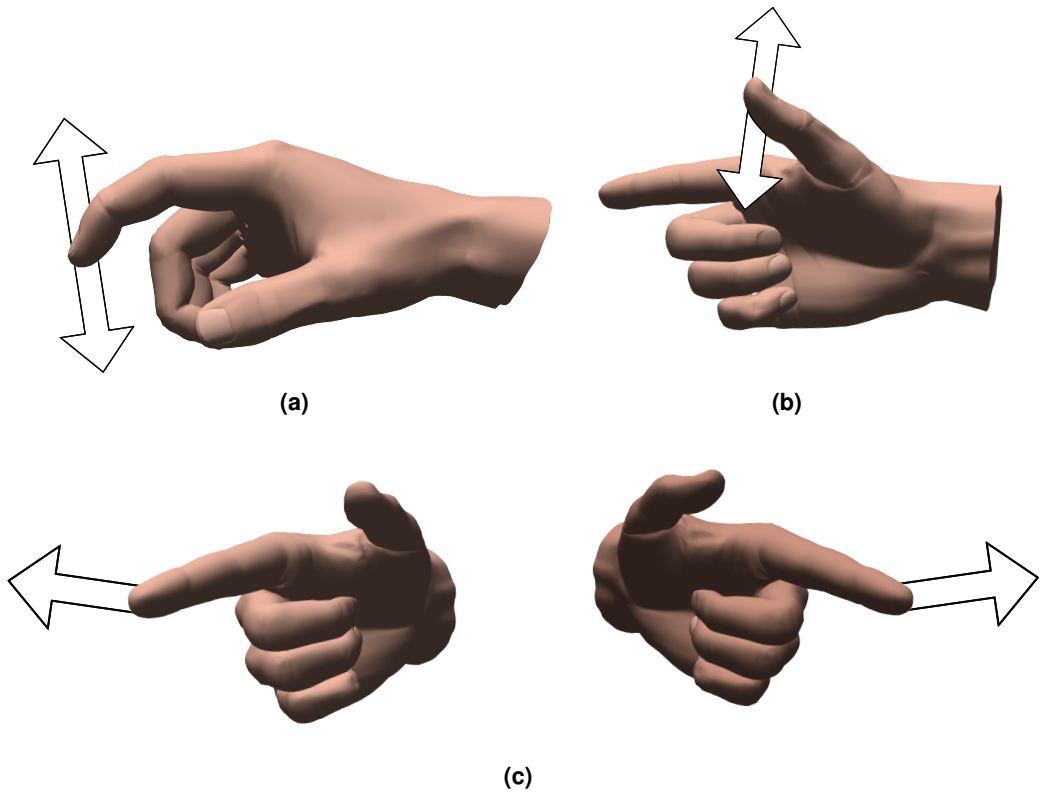


Figure 8.1: The most prevalent gestures in our studies are easy to learn and remember: *ThumbTrigger* scored best in our last experiment and it is based on a similar act compared to *AirTap*: pressing a button. Bpth (a) *AirTap* and (b) *ThumbTrigger* were preferred for selecting objects while (c) *Fingers apart* combined with *ThumbTrigger* to start and stop resizing in 2D and 3D.

8.2 Reflection

Social communities are formed, partially, by the gestures and actions that the people belonging to these communities make and are familiar with. These actions and gestures are the basis for what members consider natural. People can be part of more than one social community in their everyday lives. Examples are home and work, but also real-world and virtual communities such as online role playing games. At home we might act and gesture differently than when at work or when amongst friends. Social communities can be large, for example, vividly gesturing Italians who gesture that they do not like you so much, or they may be small, for example, underwater hockey referees who gesture that they have seen a player obstructing an opponent. Physical locations do not limit social communities, especially with the growing global village in mind [55]. You belong to a social community when, as far as the argument here is concerned, you are familiar with those actions and gestures that define that community. It is possible to become part of a social community by learning its defining actions and gestures. Examples are learning aircraft marshaling to guide aircraft or by learning the distinctive gestures of Italians [37].

Figure 8.2: The *Fingers apart* gesture as it is used on the Apple iPhone for resizing images.

Becoming part of a social community, learning its gestures, is what can explain why we found such a strong influence on our gesture set from WIMP interfaces and the iPhone interface. Our users were already part of the social community that is defined by the (inter)actions with high-tech and/or digital artifacts such as the PC. We call the members of this community 'high-tech citizens'. High-tech citizens are members of a global social community of high-tech artifact literates and users. The participants in our experiments, all of them high-tech citizens, were already familiar with WIMP actions before taking part in these experiments. An example of the actions that they were familiar with is how to delete digital objects by putting them in the virtual trash bin on the digital desktop. With the emergence of new technologies, new social communities are formed or old ones adopt these technologies. This adjusts the notion of what community members find intuitive [48]. The gesture implemented in the Apple iPhone for resizing digital objects, *Fingers apart*, based perhaps on elastic bands, is a good example of such a recent technological development. It quickly became the iPhone's most established feature since its release, racing across the planet seemingly overnight. This affected the emergence of other multi-touch technologies that were beginning to become commonplace [73]. High-tech citizens who had used an iPhone *expect* these multi-touch interfaces to have the iPhone resize gesture [44].

The iPhone's *Fingers apart* gesture has become an idea, action and perhaps even a meme [34], that is rapidly transferred from one person to the next. The people who absorb this idea and others like it then become high-tech citizens. This and other similar gestures add to the everyday manipulative actions that are the basis for the gestures that we evaluated in this thesis. In that respect, the standard WIMP paradigm has, over the past decades, indoctrinated high-tech citizens who form the potential users of the systems for which we are designing gesture-based interfaces. The WIMP paradigm is intensely familiar for these users. We demonstrate with our findings that the appeal of these existing and new interfaces provides undeniably strong reasons for preferring gestures that mimic already existing interactions. Interactions such as pressing mouse buttons and the drag-and-drop paradigm are the cause for this appeal.

It was surprising, in some way, to discover that these technological developments, some familiar such as the mouse and others more modern such as the iPhone, have such a strong impact on our findings. After all, mainstream HCI, in its everlasting search for finding the natural interface [158], focuses mainly on the influence of cultural and social settings on the interaction [159]. The purpose in this type of HCI research is to have the interaction mimic human communication or at least be based upon it for the most part [16]. The main reason for this is to make the interface as easy to use as is possible [160]. The target audience herein is as wide as possible: including the elderly, youngsters, middle-aged and teenagers alike. We believe that it merits further study to consider the extent to which there is a real need, as far as high-tech citizens are concerned, for the natural interfaces that HCI research strives after. By merit of the indoctrination by the myriad of high-tech artifacts, most potential users of the interfaces for which HCI designs interfaces might prefer direct manipulation interfaces over interface agents [191]. We readily agree that an experiment similar to ours would yield very different results in another social community whose members are high-tech illiterates. After all, they are unfamiliar with the actions and gestures that define high-tech citizens. Having realized this, HCI research is already shifting its focus on users who are not (yet) high-tech citizens. A prime example of these users might be the elderly of today that may, by introducing HCI novelties, live independently for as long as possible [95].

8.3 Future research

We first describe two practical realizations of future work that continues the line of research that has been described in this thesis. Secondly, we describe where we are now in creating an easy to use and comfortable gesture interface.

8.3.1 Practical realizations

Speech and gestures can be combined in large display interactions. In addition, our last experiment can be continued with a new wearable device, called LaserRing, with which we can compare wearable objects and bare hands for gesturing.

Deixis

One shortcoming of the gesture interfaces that we have described in this thesis is that it is not practical to input data. The traditional keyboard is much better suited for that task. However, we should prevent cases in which the user has to stop interacting and move to some input console so that some requested data input can take place. One well-studied solution for this is to combine speech and gesturing. Speech is the prevalent modality that is used in conjunction with gesturing for building multi-modal large display interfaces [174]. It is hard to build these multi-modal interfaces [197]: due to the nature of spontaneous, unplanned and unconsciously made gestures, it is hard to recognize and interpret their meaning without the speech that they enrich [133; 195]. As a result, gestures have been included in human-computer interfaces mainly to serve as a means to disambiguate other input modalities, for example, speech [222]. Our findings have shown that users explicitly mark the beginning and ending of their gestures by tensing and relaxing their hands. This cue can be used to detect gesture segmentation from continuous gesturing [75].

Huijbregts [86] describes a large vocabulary continuous speech recognition system called SHoUT that was developed at our research group. This system can be used to robustly detect, in real-time, speech from non-speech, how many speakers are speaking and what they are saying with good accuracy. Matching these two modalities, gesture and speech, is hard to do because co-verbal gesturing is very user-dependent [118]. Van der Sluis [208] and Oviatt [157] describe pattern matching techniques that can detect a user's prevalent input pattern of speech and gesturing so that this pattern can describe matches between gesture and speech. Both gesturing and speech can be used to detect the beginning and ending of the interaction with the display. It is often not possible to detect these events, making it difficult to reach Buxton's null-state in which the user is out-of-range of the interaction [19].

LaserRing

The gloves that we used in our last experiment, see Chapter 7, were a first prototype. Since then, we have experimented with smaller batteries, buttons directly attached to the laser and the whole device sewn onto a cotton glove. Our users strongly indicated, in the evaluation of the prototype glove, that they prefer minimal gestures to issue commands. No users mentioned that they disliked wearing the glove to interact with the display. In fact, they found it comfortable to interact in that manner. This leads us to believe that our prototype gloves might be miniaturized to be less invasive and still provide an explicit means to interact with large displays at a distance. We have already put forth the idea of using infrared lasers so that the smoothness and accuracy of the device can be increased because we suspect that it will benefit the interaction, as was also argued by König *et al.* [114]. The miniaturized version of our prototype will be called 'LaserRing'. The device will still be mounted on the back of the hand so that the laser can be pointed accurately without being affected by the user moving his fingers [12]. Buttons will still be used

to issue commands and they will be placed on the prevalent spots on the fingers that we observed in our last experiment.

We propose to use the LaserRings to investigate to what extent there is a preference for using a wearable device to interact with a large display at a distance or to not use a device at all. Tangible objects provide the user with an explicit means to interact. In addition, objects that a user picks up signal to other users whose turn it is. Wearable devices might not have quite the same result but we can find out what the difference is by comparing a handheld device, for example, a Wiimote [46], the LaserRing and bare hands [132]. Without a wearable device to gesture with, the interaction is based on unobtrusive means to look at the user, for example, through cameras. The immature techniques for hand shape recognition possibly hamper this comparison [187]. To overcome this issue, we might focus on detecting the hole formed by performing *ThumbTrigger* rather than detecting the hand shape [229] or we might turn to the use of colored gloves to assist the computer vision recognition process [100].

Another aspect that we propose to use LaserRings for is to investigate the user experience during prolonged interactions [199, Ch.2]. Physical fatigue during these prolonged interactions is often mentioned as a possible downside in human-computer interactions [94] and it can degrade performance [5]. Cerney and Vance [25] propose the use of quick and easy gestures, as we used in this thesis, to reduce fatigue and favor the ease of learning. We suspect that adding another, increasingly popular and widespread technology in the mix will reduce the users perception of fatigue even more. The most fatiguing aspect in gesture interfaces is the lack of a means to rest the arms and hands. We propose to use multi-touch sensitive surfaces, introduced by Han [73], as a physical surface on which the hands can be rested. This surface might be horizontal, titled or perhaps even vertical. By combining the LaserRings for gesturing at a distance and multi-touch surfaces for interacting for prolonged durations with ease, comfort, without causing (physical) fatigue and preventing 'gorilla-arms'.

8.3.2 Where are we now?

The focus of this thesis has been on gesture interactions at a distance. The guideline for designing gesture sets for these interfaces, see Section 8.1, is that the gestures should be movements that are, above all, simple and minimal. This is, as far was we are concerned, what makes the gestures that John Anderton makes in Minority Report so believable. So, how far are we from building the gesture interface that John uses, or others like it? As we have shown in this thesis, there have already been numerous attempts to build easy to learn gesture interfaces: not in the last place the ongoing attempt by the people involved with creating Minority Report in the first place[1] to commercialize a similar gesture interface. The most immature piece of easy to use and comfortable gesture interfaces is no longer the understanding of the gesture interaction itself. It is the availability and accessibility of the sensors

[1] MIT spin-off company Oblong Industries, `http://oblong.com`, June 16th, 2009.

that look at the user that limit the widespread construction of gesture interfaces.

We predict that when sensors become available and accessible on a large scale for low costs, the speed of developments and the number of developers increase exponentially. The Nintendo Wiimote game controller is one recent example. Its hardware has been taken apart by hundreds and hundreds of enthusiasts while numerous open source reverse engineering initiatives developed the software needed to gain access to the Wiimote's sensors. Based on those developments, these enthusiasts have been using the cheap sensors in the Wiimote for all kinds of things: from head tracking to creating affordable digital white boards for schools [123]. Another example is found in the emergence of multi-touch technology over the past couple of years. Existing multi-touch technologies, for example, the DiamondTouch [35], were expensive and prevented fast growth of both the understanding and use of multi-touch interfaces [44]. Han [73] showed how easy and cheap it is to create your own multi-touch interface and promptly most computer science departments across the globe had at least one student who was building his own multi-touch table. Various open source software initiatives, for example, the NUI group[2], have developed means to easily gain access to the user's touches that have resulted in music tables [45; 98], games [44; 204] and collaborative design applications [206; 236]. A recently announced, cheap sensor solution is Microsoft's Project Natal[3] that will be equipped with a multi-array microphone, RGB camera and a depth sensor. We strongly believe that just as with the Wiimote, this sensor will be embraced by similar developers so that entirely new applications will be made possible.

The fun of exploring the workings and capabilities of easily accessible, cheap sensors is what seemingly triggers such responses. It will be worthwhile to find means to make sensors more available and accessible for exploratory research in addition to hardware and software developers. We believe that then a revolution can take place that can make gesture interfaces even more common place as the multi-touch revolution has achieved so far.

[2] `http://www.nuigroup.com/`, November 17th, 2009.

[3] `http://www.xbox.com/en-US/live/projectnatal/`, November 17th, 2009.

Bibliography

[1] A. Agarawala and R. Balakrishnan. Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI '06)*, pp. 1283–1292. ACM, New York, NY, USA: 2006.

[2] D. Ahlstroem, R. Alexandrowicz, and M. Hitz. Improving menu interaction: a comparison of standard, force enhanced and jumping menus. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI '06)*, pp. 1067–1076. ACM, New York, NY, USA: 2006.

[3] C. Appert and S. Zhai. Using strokes as command shortcuts: cognitive benefits and toolkit support. In: *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*, pp. 2289–2298. ACM, New York, NY, USA: 2009.

[4] J. Bardram, C. Bossen, A. Lykke-Olesen, R. Nielsen, and K. H. Madsen. Virtual video prototyping of pervasive healthcare systems. In: *Proceedings of the 4th conference on Designing interactive systems (DIS '02)*, pp. 167–177. ACM, New York, NY, USA: 2002.

[5] T. Baudel and M. Beaudouin-Lafon. Charade: remote control of objects using free-hand gestures. *Communications of the ACM*, **vol. 36 (7)**: 28–35: 1993.

[6] T. Beauvisage. Computer usage in daily life. In: *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*, pp. 575–584. ACM, New York, NY, USA: 2009.

[7] F. Bérard, J. Ip, M. Benovoy, D. El-Shimy, J. Blum, and J. Cooperstock. Did "Minority Report" Get It Wrong? Superiority of the Mouse over 3D Input Devices in a 3D Placement Task. In: *Human-Computer Interaction (INTERACT '09)*, vol. 5727/2009 of *Lecture Notes in Computer Science*, pp. 400–414. Springer Berlin / Heidelberg: 2009.

[8] N. Beringer. Evoking Gestures in SmartKom - Design of the Graphical User Interface. In: *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915/2004 of *Lecture Notes in Computer Science*, pp. 409–420. Springer Berlin / Heidelberg: 2002.

[9] H.-J. Bieg. *Laserpointer and Eye Gaze Interaction - Design and Evaluation*. mastersthesis, University of Konstanz: 2008.

[10] K. Boffard. Abdominal trauma. In: *Core Topics in General and Emergency Surgery*, vol. 1 of *Companion to specialist surgical practice series*, chap. 13, pp. 239–260. Elsevier Health Sciences, 3 ed.: 2006.

[11] R. Bolt. "Put-that-there": Voice and gesture at the graphics interface. *SIGGRAPH Computer Graphics*, **vol. 14 (3)**: 262–270: 1980.

[12] B. Bongers. *Interaction with our electronic environment – an e-cological approach to physical interface design*, vol. 34 of *Cahier*. Faculty of Journalism and Communication, Hogeschool van Utrecht: 2004.

[13] D. Bowman, E. Kruijff, J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*.

Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA: 2004.

[14] A. Bragdon, R. Zeleznik, B. Williamson, T. Miller, and J. LaViola. GestureBar: improving the approachability of gesture-based interfaces. In: *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*, pp. 2269–2278. ACM, New York, NY, USA: 2009.

[15] H. Brignull and Y. Rogers. Enticing People to Interact with Large Public Displays in Public Spaces. In: *Proceedings of the IFIP International Conference on Human-Computer Interaction (INTERACT '03)*, pp. 17–24: 2003.

[16] H. Bunt, M. Kipp, M. Maybury, and W. Wahlster. Fusion and coordination for multimodal interactive information presentation. In: W. Verhaegh, E. Aarts, and J. Korst, eds., *Algortihms in Ambient Intelligence*, vol. 2 of *Philips Research Book Series*, pp. 21–53. Kluwer Academic Publishers, Boston/Dordrecht/London: 2003.

[17] M. Büscher, S. Gill, P. Mogensen, and D. Shapiro. Landscapes of Practice: Bricolage as a Method for Situated Design. *Computer Supported Cooperative Work (CSCW)*, **vol. 10 (1)**: 1–28: 2001.

[18] B. Butterworth and U. Hadar. Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, **vol. 96 (1)**: 168–174: 1989.

[19] W. Buxton. A three-state model of graphical input. In: *Proceedings of the IFIP TC13 Third Interational Conference on Human-Computer Interaction (INTERACT '90)*, pp. 449–456. North-Holland Publishing Co., Amsterdam, The Netherlands: 1990.

[20] W. Buxton. Chunking and Phrasing and the Design of Human-Computer Dialogues. In: *Human-computer interaction: toward the year 2000*, pp. 494 – 499. Morgan Kaufmann Publishers Inc.: 1995.

[21] S. Card, A. Newell, and T. Moran. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA: 1983.

[22] J. Cassell. A Framework For Gesture Generation And Interpretation. In: *Computer Vision in Human-Machine Interaction*, pp. 191–215. Cambridge University Press: 1998.

[23] J. Cassell, D. McNeill, and K.-E. McCullough. Speech-Gesture Mismatches: Evidence for One Underlying Representation of Linguistic and Non-Linguistic Information. *Pragmatics and Cognition*, **vol. 7 (1)**: 1–33: 1998.

[24] Á. Cassinelli, S. Perrin, and M. Ishikawa. Smart laser-scanner for 3D human-machine interface. In: *Extended abstracts on Human factors in computing systems (CHI '05)*, pp. 1138–1139. ACM, New York, NY, USA: 2005.

[25] M. Cerney and J. Vance. Gesture Recognition in Virtual Environments: A Review and Framework for Future Developments. *Technical Report ISU-HCI-2005-01*, Iowa State University Human Computer Interaction, Ames, Iowa: 2005.

[26] A. Chan, R. Lau, and L. Li. Hand Motion Prediction for Distributed Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, **vol. 14 (1)**: 146–159: 2008.

[27] M. Chen. A framework for describing interactions with graphical widgets. In: *INTERACT '93 and CHI '93 conference companion on Human factors in computing systems (CHI '93)*, pp. 131–132. ACM, New York, NY, USA: 1993.

[28] K. Cheng and M. Takatsuka. Hand Pointing Accuracy for Vision-Based Interactive Systems. In: *Human-Computer Interaction (INTERACT '09)*, vol. 5727/2009 of *Lecture Notes in Computer Science*, pp. 13–16. Springer Berlin / Heidelberg: 2009.

[29] R. Clark. Media are “mere vehicles” – the opening argument. In: *Learning from media – arguments, analysis, and evidence*, vol. 1 of *Perspectives in Instructional Technology and Distance Learning*, pp. 1–12. Information Age Publishing: 2001.

[30] C. Cohen. A Brief Overview of Gesture Recognition. online: 1999.

[31] L. Cutler, B. Froehlich, and P. Hanrahan. Two-handed direct manipulation on the responsive workbench. In: *Proceedings of the 1997 symposium on Interactive 3D graphics (SI3D ’97)*, pp. 107–114. ACM Press, New York, NY, USA: 1997.

[32] M. Czerwinski, G. Robertson, B. Meyers, G. Smith, D. Robbins, and D. Tan. Large display research overview. In: *Extended abstracts on Human factors in computing systems (CHI ’06)*, pp. 69–74. ACM Press, New York, NY, USA: 2006.

[33] R. B. D’Agostino and M. A. Stephens. *Goodness-of-fit Techniques*, vol. 68 of *Statistics: textbooks and monographs*. CRC Press: 1986.

[34] R. Dawkins. *The Selfish Gene*. Oxford University Press: 1990.

[35] P. Dietz and D. Leigh. DiamondTouch: a multi-user touch technology. In: *Proceedings of the 14th annual ACM symposium on User interface software and technology (UIST ’01)*, pp. 219–226. ACM Press, New York, NY, USA: 2001.

[36] A. Dix, M. Ghazali, and D. Ramduny-Ellis. Modelling Devices for Natural Interaction. *Electronic Notes in Theoretical Computer Science*, **vol. 208**: 23 – 40: 2008.

[37] D. Efron. *Gesture and Environment*. King’s Crown Press, New York: 1941.

[38] D. Efron. *Gesture, race and culture; a tentative study of the spatio-temporal and "linguistic" aspects of the gestural behavior of eastern Jews and southern Italians in New York City, living under similar as well as different environmental conditions.* Approaches to semiotics, 9. Mouton, The Hague: 1972.

[39] P. Ekman and W. Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, **vol. 1**: 49–98: 1969.

[40] J. Epps, S. Lichman, and M. Wu. A study of hand shape use in tabletop gesture interaction. In: *Extended abstracts on Human factors in computing systems (CHI ’06)*, pp. 748–753. ACM, New York, NY, USA: 2006.

[41] A. Erol, G. Bebis, M. Nicolescu, R. Boyle, and X. Twombly. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, **vol. 108 (1-2)**: 52–73: 2007.

[42] R. Fiebrink, D. Morris, and M. Morris. Dynamic mapping of physical controls for tabletop groupware. In: *Proceedings of the 27th international conference on Human factors in computing systems (CHI ’09)*, pp. 471–480. ACM, New York, NY, USA: 2009.

[43] W. Fikkert, M. D’Ambros, T. Bierz, and T. Jankun-Kelly. Interacting with Visualizations. In: A. Kerren, A. Ebert, and J. Meyer, eds., *Human-Centered Visualization Environments*, vol. 4417/2007 of *Lecture Notes in Computer Science*, GI-Dagstuhl Seminar 3, pp. 77–162. Springer Verlag: 2007.

[44] W. Fikkert, M. Hakvoort, P. van der Vet, and A. Nijholt. Experiences with interactive multi-touch tables. In: *The 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN ’09)*, vol. 9 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 193–200. Springer Berlin Heidelberg: 2009.

[45] W. Fikkert, M. Hakvoort, P. van der Vet, and A. Nijholt. FeelSound: Collaborative Composing of Acoustic Music. In: *Proceedings of the 6th International Conference on Advances in Computer*

Entertainment Technology (ACE '09), pp. 294–297. ACM, Athens, Greece: 2009.

[46] W. Fikkert, N. Hoeijmakers, P. van der Vet, and A. Nijholt. Navigating a Maze with Balance Board and Wiimote. In: *The 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN '09)*, vol. 9 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 187–192: 2009.

[47] W. Fikkert, H. van der Kooij, Z. Ruttkay, and H. van Welbergen. Measuring Behavior using Motion Capture. In: A. Spink, M. Ballintijn, N. Bogers, F. Grieco, L. Loijens, L. Noldus, G. Smit, and P. Zimmerman, eds., *Proceedings of Measuring Behavior 2008, 6th International Conference on Methods and Techniques in Behavioral Research*, pp. 13–13. Noldus, Maastricht, The Netherlands: 2008.

[48] W. Fikkert, P. van der Vet, and A. Nijholt. Gestures for Large Display Control. In: *Gesture in Embodied Communication and Human-Computer Interaction*, vol. 5934/2009 of *Lecture Notes in Computer Science*, p. 12. Springer, Berlin: 2009.

[49] W. Fikkert, P. van der Vet, H. Rauwerda, T. Breit, and A. Nijholt. A Natural Gesture Repertoire for Cooperative Large Display Interaction. In: *Advances in Gesture-Based Human-Computer Interaction and Simulation*, vol. 5085/2009 of *Lecture Notes on Computer Science*, chap. 22, pp. 199–204. Springer Berlin / Heidelberg: 2009.

[50] A. Fisch, C. Mavroidis, J. Melli-Huber, and Y. Bar-Cohen. Chapter 4: Haptic Devices for Virtual Reality, Telepresence, and Human-Assistive Robotics. In: Y. Bar-Cohen and C. Breazeal, eds., *Biologically-Inspired Intelligent Robots*. SPIE Press: 2003.

[51] C. Forlines and R. Lilien. Adapting a single-user, single-display molecular visualization application for use in a multi-user, multi-display environment. In: *Proceedings of the working conference on Advanced visual interfaces (AVI '08)*, pp. 367–371. ACM, New York, NY, USA: 2008.

[52] C. Forlines, D. Wigdor, C. Shen, and R. Balakrishnan. Direct-touch vs. mouse input for tabletop displays. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07)*, pp. 647–656. ACM Press, New York, NY, USA: 2007.

[53] D. Forsyth and J. Ponce. *Computer Vision - A modern approach*. Pearson Education, Inc.: 2003.

[54] N. Freedman and S. P. Hoffman. Kinetic behavior in altered clinical states: approach to objective analysis of motor behavior during clinical interviews. *Perceptual and motor skills*, **vol. 24 (2)**: 527–539: 1967.

[55] T. Friedman. *The World Is Flat: A Brief History of the Twenty-first Century*. Farrar, Straus and Giroux: 2005.

[56] B. Frohlich, J. Plate, J. Wind, G. Wesche, and M. Gobel. Cubic-Mouse-based interaction in virtual environments. *IEEE Computer Graphics and Applications*, **vol. 20 (4)**: 12–15: 2000.

[57] D. Gavrila. The visual analysis of human movement: a survey. *Computer Vision and Image Understanding*, **vol. 73 (1)**: 82–98: 1999.

[58] S. Gibet, N. Courty, and J.-F. Kamp, eds. *Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers*, vol. 3881 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg: 2006.

[59] S. Gibet and P.-F. Marteau. Analysis of Human Motion, Based on the Reduction of Multidimensional Captured Data — Application to Hand Gesture Compression, Segmentation and Synthesis. In: *Proceedings of the 5th international conference on Articulated Motion and Deformable Objects (AMDO '08)*, pp. 72–81. Springer-Verlag, Berlin, Heidelberg: 2008.

[60] S. Gill and J. Borchers. Knowledge in co-action: social intelligence in collaborative design activity. *AI and Society*, **vol. 17 (3)**: 322–339: 2003.

[61] G. Goggin. Adapting the mobile phone: The iPhone and its consumption. *Continuum*, **vol. 23 (2)**: 231–244: 2009.

[62] A. Gonzales, T. Finley, and S. Duncan. (Perceived) interactivity: does interactivity increase enjoyment and creative identity in artistic spaces? In: *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*, pp. 415–418. ACM, New York, NY, USA: 2009.

[63] T. Grossman, R. Balakrishnan, G. Kurtenbach, G. Fitzmaurice, A. Khan, and W. Buxton. Creating principal 3D curves with digital tape drawing. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '02)*, pp. 121–128. ACM Press, New York, NY, USA: 2002.

[64] T. Grossman, D. Wigdor, and R. Balakrishnan. Multi-finger gestural interaction with 3d volumetric displays. In: *Proceedings of the 17th annual ACM symposium on User interface software and technology (UIST '04)*, pp. 61–70. ACM Press, New York, NY, USA: 2004.

[65] Y. Guiard. Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. *Journal of Motor Behavior*, **vol. 19 (4)**: 486–517: 1987. Slightly edited version of an article originally published.

[66] Y. Guiard. On Fitts's and Hooke's laws: simple harmonic movement in upper-limb cyclical aiming. *Acta psychologica*, **vol. 82 (1-3)**: 139–159: 1993.

[67] Y. Guiard and M. Beaudouin-Lafon. Target acquisition in multiscale electronic worlds. *International Journal of Human-Computer Studies*, **vol. 61 (6)**: 875–905: 2004.

[68] Y. Guiard, M. Beaudouin-Lafon, and D. Mottet. Navigation as multiscale pointing: extending Fitts' model to very high precision tasks. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '99)*, pp. 450–457. ACM Press, New York, NY, USA: 1999.

[69] M. Gullberg, H. Hendriks, and M. Hickmann. Learning to talk and gesture about motion in French. *First Language*, **vol. 28 (2)**: 200–236: 2008.

[70] M. Hachet, J. Pouderoux, and P. Guitton. A camera-based interface for interaction with mobile handheld computers. In: *Proceedings of the 2005 symposium on Interactive 3D graphics and games (SI3D '05)*, pp. 65–72. ACM Press, New York, NY, USA: 2005.

[71] M. Hafez. Tactile interfaces: technologies, applications and challenges. *The Visual Computer*, **vol. 23 (4)**: 267–272: 2004.

[72] E. Hall. *The Hidden Dimension: Man's Use of Space in Public and Private*. New York: Doubleday, 2 ed.: 1990.

[73] J. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In: *Proceedings of the 18th annual ACM symposium on User interface software and technology (UIST '05)*, pp. 115–118. ACM Press, New York, NY, USA: 2005.

[74] D. Hansen. *Committing Eye Tracking*. Ph.D. thesis, IT University of Copenhagen: 2003.

[75] P. Harling and A. Edwards. Hand Tension as a Gesture Segmentation Cue. In: *Gesture Workshop*, pp. 75–88. Springer: 1996.

[76] C. Harrison and S. Hudson. Providing dynamically changeable physical buttons on a visual display. In: *Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI '09)*, pp. 299–308. ACM, New York, NY, USA: 2009. Session: Clicking on buttons.

[77] A. Hauptmann. Speech and gestures for graphic image manipulation. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '89)*, pp. 241–245. ACM Press, New York, NY, USA: 1989.

[78] A. Heuser, H. Kourtev, S. Winter, D. Fensterheim, G. Burdea, V. Hentz, and P. Forducey. Tele-Rehabilitation using the Rutgers Master II glove following Carpal Tunnel Release surgery. In: *International Workshop on Virtual Rehabilitation*, pp. 88–93: 2006.

[79] K. Hinckley. Input technologies and techniques. In: A. Sears and J. Jacko, eds., *Handbook of Human-Computer Interaction: fundamentals, evolving technologies and emerging applications*, chap. 7, pp. 151–168. Lawrence Erlbaum Associates Inc., Hillsdale, NJ, USA: 2006.

[80] K. Hinckley, M. Czerwinski, and M. Sinclair. Interaction and modeling techniques for desktop two-handed input. In: *Proceedings of the 11th annual ACM symposium on User interface software and technology (UIST '98)*, pp. 49–58. ACM Press, New York, NY, USA: 1998.

[81] K. Hinckley, R. Pausch, D. Proffitt, and N. Kassell. Two-handed virtual manipulation. *ACM Transactions on Computer-Human Interaction*, **vol. 5 (3)**: 260–302: 1998.

[82] K. Hinckley, G. Ramos, F. Guimbretière, P. Baudisch, and M. Smith. Stitching: pen gestures that span multiple displays. In: *Proceedings of the working conference on Advanced visual interfaces (AVI '04)*, pp. 23–31. ACM Press, New York, NY, USA: 2004.

[83] K. Hinckley, J. Tullio, R. Pausch, D. Proffitt, and N. Kassell. Usability analysis of 3D rotation techniques. In: *Proceedings of the 10th annual ACM symposium on User interface software and technology (UIST '97)*, pp. 1–10. ACM, New York, NY, USA: 1997.

[84] D. Holman, R. Vertegaal, M. Altosaar, N. Troje, and D. Johns. Paper windows: interaction techniques for digital paper. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '05)*, pp. 591–599. ACM Press, New York, NY, USA: 2005.

[85] S. Hotelling and B. Huppi. Force imaging input device and system. *Technical Report 7.538.760*, US Patent: 2006.

[86] M. Huijbregts. *Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled*. Ph.D. thesis, University of Twente, Enschede: 2008.

[87] C. Hummels, G. Smets, and K. Overbeeke. An Intuitive Two-Handed Gestural Interface for Computer Supported Product Design. In: *Gesture and Sign Language in Human-Computer Interaction*, vol. 1371/1998 of *Lecture Notes in Computer Science*, p. 197. Springer Berlin / Heidelberg: 1998.

[88] C. Hummels and P. J. Stappers. Meaningful Gestures for Human Computer Interaction: Beyond Hand Postures. In: *Proceedings of the 3rd. International Conference on Face and Gesture Recognition (FG '98)*, vol. 3 of *Automatic Face and Gesture Recognition*, pp. 591–596. IEEE Computer Society, Washington, DC, USA: 1998.

[89] I. Incertis, J. Garcia-Bermejo, and E. Casanova. Hand Gesture Recognition for Deaf People Interfacing. In: *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, pp. 100–103. IEEE Computer Society, Washington, DC, USA: 2006.

[90] K. Ishii, S. Zhao, M. Inami, T. Igarashi, and M. Imai. Designing Laser Gesture Interface for Robot Control. In: *Human-Computer Interaction (INTERACT 2009)*, vol. 5727/2009 of *Lecture Notes in Computer Science*, pp. 479–492. Springer Berlin / Heidelberg: 2009.

[91] ISO. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices. *Technical Report ISO/DIS 9241-9:2000*, International Organization for Standardization: 2000.

[92] P. Isokoski, R. Raisamo, B. Martin, and G. Evreinov. User performance with trackball-mice.

Interacting with Computers, **vol. 19 (3)**: 407–427: 2007.

[93] A. Jaimes and N. Sebe. Multimodal Human Computer Interaction: A Survey. *Computer Vision and Image Understanding*, **vol. 108 (1-2)**: 116–134: 2007.

[94] B. Jensen, B. Laursen, and A. Ratkevicius. Forearm Muscular Fatigue during four Hours of Intensive Computer Mouse Work - Relation to Age. In: *Proceedings of HCI International (the 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I*, pp. 93–96. L. Erlbaum Associates Inc., Hillsdale, NJ, USA: 1999.

[95] J. Jorge. Adaptive tools for the elderly: new devices to cope with age-induced cognitive disabilities. In: *Proceedings of the 2001 EC/NSF workshop on Universal accessibility of ubiquitous computing (WUAUC '01)*, pp. 66–70. ACM, New York, NY, USA: 2001.

[96] K. Kahol, P. Tripathi, and S. Panchanathan. Automated gesture segmentation from dance sequences. In: *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition (FG '04)*, pp. 883–888: 2004.

[97] M. Kaltenbrunner, T. Bovermann, R. Bencina, and E. Costanza. TUIO: A Protocol for Table-Top Tangible User Interfaces. In: *Proceedings of the 6th International Workshop on Gesture in Human-Computer Interaction and Simulation (GW 2005)*: 2005.

[98] M. Kaltenbrunner, S. Jorda, G. Geiger, and M. Alonso. The reacTable*: A Collaborative Musical Instrument. In: *Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE '06)*, pp. 406–411. IEEE Computer Society, Washington, DC, USA: 2006.

[99] T. Kanno, K. Nakata, and K. Furuta. A method for conflict detection based on team intention inference. *Interacting with Computers*, **vol. 18 (4)**: 747–769: 2006.

[100] M. Karam, J. Hare, and M. Schraefel. Ambient Gestures. *Technical Report ECSTR-IAM06-001*, Intelligence, Agents, Multimedia Group, School of Electronics and Computer Science, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom: 2004.

[101] M. Karam and M. Schraefel. A study on the use of semaphoric gestures to support secondary task interactions. In: *Extended abstracts on Human factors in computing systems (CHI '05)*, pp. 1961–1964. ACM, New York, NY, USA: 2005.

[102] M. Karam and M. Schraefel. A Taxonomy of Gestures in Human Computer Interactions. *Technical Report ECSTR-IAM05-009*, Electronics and Computer Science, University of Southampton: 2005.

[103] M. Kavakli and D. Jayarathna. Virtual Hand: An Interface for Interactive Sketching in Virtual Reality. *International Conference on Computational Intelligence for Modelling, Control and Automation*, **vol. 1**: 613–618: 2005.

[104] M. Kavakli, M. Taylor, and A. Trapeznikov. Designing in virtual reality (DesIRe): a gesture-based interface. In: *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts (DIMEA '07)*, pp. 131–136. ACM, New York, NY, USA: 2007.

[105] A. Kendon. Some relationships between body motion and speech. In: A. Siegman and B. Pope, eds., *Studies in dyadic communication*, vol. 7 of *General Psychology Series*, pp. 177–210. Pergamon Press: 1972.

[106] A. Kendon. Gesticulation and speech: two aspects of the process of utterance. In: R. Key, ed., *Relationship of Verbal and Nonverbal Communication*, pp. 207–228. Mouton: 1980.

[107] A. Kendon. Current issues in the study of gesture. *Journal for the anthropological study of human movement*, **vol. 5 (3)**: 101–134: 1989. Reprinted.

[108] A. Kendon. *Gesture: Visible Action as Utterance*. Cambridge University Press: 2004.

[109] L. Kim and S. H. Park. A Haptic Sculpting Technique Based on Volumetric Representation. In: *Articulated Motion and Deformable Objects*, vol. 3179/2004 of *Lecture Notes in Computer Science*, pp. 14–25. Springer Berlin / Heidelberg: 2004.

[110] M. Kipp. *Gesture Generation by Imitation - From Human Behavior to Computer Character Animation*. Ph.D. thesis, Saarland University, Saarbruecken, Germany, Boca Raton, Florida: 2004.

[111] M. Kipp, M. Neff, and I. Albrecht. An Annotation Scheme for Conversational Gestures : How to economically capture timing and form. In: *Proceedings of the Workshop on Multimodal Corpora at LREC 2006*, pp. 24–27: 2006.

[112] D. Kirk, T. Rodden, and D. Stanton Fraser. Turn it this way: grounding collaborative action with remote gestures. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07)*, pp. 1039–1048. ACM, New York, NY, USA: 2007.

[113] W. König, H.-J. Bieg, and H. Reiterer. Laserpointer - Interaktion für große, hochauflösende Displays. In: *Mensch und Computer 2007: Interaktion im Plural, 7. Konferenz für interaktive und kooperative Medien*, pp. 69–78. Oldenbourg Verlag: 2007.

[114] W. König, J. Böttger, N. Völzow, and H. Reiterer. Laserpointer-Interaction between Art and Science. In: *Proceedings of the 13th international conference on Intelligent User Interfaces (IUI'08)*, pp. 423–424. ACM Press: 2008.

[115] S. Kopp and I. Wachsmuth. Synthesizing multimodal utterances for conversational agents: Research Articles. *Computer Animation and Virtual Worlds*, **vol. 15 (1)**: 39–52: 2004.

[116] O. Korner and R. Manner. Implementation of a haptic interface for a virtual reality simulator for flexible endoscopy. In: *Proceedings of the 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS '03)*, pp. 278–284: 2003.

[117] K. J. Kortbek and K. Grønbæk. Interactive spatial multimedia for communication of art in the physical museum space. In: *Proceedings of the 16th ACM international conference on Multimedia (MM '08)*, pp. 609–618. ACM, New York, NY, USA: 2008.

[118] A. Kranstedt, P. Kühnlein, and I. Wachsmuth. Deixis in Multimodal Human Computer Interaction: An Interdisciplinary Approach. In: *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915/2004 of *Lecture Notes in Computer Science*, chap. 11, pp. 112–123. Springer Berlin, Heidelberg: 2004.

[119] A. Kranstedt, A. Lücking, T. Pfeiffer, H. Rieser, and I. Wachsmuth. Deixis: How to Determine Demonstrated Objects Using a Pointing Cone. In: [58], chap. 34, pp. 300–311.

[120] R. Krauss, Y. Chen, and R. Gottesman. Lexical Gestures and Lexical Access: A Process Model. In: D. McNeill, ed., *Language and gesture*, chap. 13, pp. 261–283. Cambridge University Press, New York: 2000.

[121] A. Kron and G. Schmidt. Haptisches Telepräsenzsystem zur Unterstützung bei Entschärfungstätigkeiten: Systemgestaltung, Regelung und Evaluation (Haptic Telepresence System for Support of Disposal of Explosive Ordnances: Design Issues, Control, and Evaluation). *at - Automatisierungstechnik*, **vol. 53 (3)**: 101–113: 2005.

[122] H.-K. Lee and J.-H. Kim. Gesture spotting from continuous hand motion. *Pattern Recognition Letters*, **vol. 19 (5-6)**: 513–520: 1998.

[123] J. C. Lee. Hacking the Nintendo Wii Remote. *IEEE Pervasive Computing*, **vol. 7 (3)**: 39–45: 2008.

[124] S. U. Lee and I. Cohen. 3D Hand Reconstruction from a Monocular View. In: *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 3, pp. 310–313. IEEE Computer Society, Washington, DC, USA: 2004.

[125] J. Lester, S. Converse, S. Kahler, T. Barlow, B. Stone, and R. Bhoga. The Persona Effect: Affective Impact of Animated Pedagogical Agents. In: *CHI '97 extended abstracts on Human factors in computing systems*, pp. 359–366. ACM Press, Atlanta, USA: 1997.

[126] J. Lichtenauer, G. ten Holt, E. Hendriks, and M. Reinders. Sign language detection using 3D visual cues. In: *IEEE Conference on Advanced Video and Signal Based Surveillance (AVVS'07)*, pp. 435–440: 2007.

[127] M. Lyons. Facial gesture interfaces for expression and communication. In: *IEEE International Conference on Systems, Man and Cybernetics*, vol. 1, pp. 598–603: 2004.

[128] S. MacKenzie. *Fitts' Law as a Performance Model in Human-Computer Interaction*. Ph.D. thesis, University of Toronto, Toronto, Ontario, Canada: 1991.

[129] S. MacKenzie, X. Zhang, and W. Soukoreff. Text entry using soft keyboards. *Behaviour and Information Technology*, **vol. 18 (4)**: 235–244: 1999.

[130] P. Majaranta, U.-K. Ahola, and O. Špakov. Fast gaze typing with an adjustable dwell time. In: *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*, pp. 357–360. ACM, New York, NY, USA: 2009.

[131] S. Malik and J. Laszlo. Visual touchpad: a two-handed gestural input device. In: *Proceedings of the 6th international conference on Multimodal interfaces (ICMI '04)*, pp. 289–296. ACM Press, New York, NY, USA: 2004.

[132] S. Malik, A. Ranjan, and R. Balakrishnan. Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In: *Proceedings of the 18th annual ACM symposium on User interface software and technology (UIST '05)*, pp. 43–52. ACM Press, New York, NY, USA: 2005.

[133] S. McKenna, G. McAllister, and I. Ricketts. Hand tracking for behaviour understanding. *Image and Vision Computing*, **vol. 20 (12)**: 827–840: 2002.

[134] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, Chicago: 1992.

[135] P. Mistry, P. Maes, and L. Chang. WUW - wear Ur world: a wearable gestural interface. In: *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (CHI EA '09)*, pp. 4111–4116. ACM, New York, NY, USA: 2009.

[136] T. Moeslund, A. Hilton, and V. Kruger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, **vol. 104 (2-3)**: 90–126: 2006.

[137] T. Moeslund, M. Störring, and E. Granum. A Natural Interface to a Virtual Environment through Computer Vision-Estimated Pointing Gestures. In: *Gesture and Sign Language in Human-Computer Interaction*, vol. 2298/2002 of *Lecture Notes in Computer Science*, chap. 6, pp. 239–250. Springer Berlin / Heidelberg: 2002.

[138] M. Morris, A. Huang, A. Paepcke, and T. Winograd. Cooperative gestures: multi-user gestural interactions for co-located groupware. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI '06)*, pp. 1201–1210. ACM, New York, NY, USA: 2006.

[139] O. Mubin, T. Lashina, and E. van Loenen. How Not to Become a Buffoon in Front of a Shop Window: A Solution Allowing Natural Head Movement for Interaction with a Public Display. In: *Human-Computer Interaction (INTERACT '09)*, vol. 5727/2009 of *Lecture Notes in*

Computer Science, pp. 250–263. Springer Berlin / Heidelberg: 2009.

[140] J. Mulder and R. van Liere. The Personal Space Station: Bringing Interaction Within Reach. In: *Proceedings of the 4th Virtual Reality International Conference (VRIC '02)*, pp. 73–81. IEEE: 2002.

[141] B. Myers, R. Bhatnagar, J. Nichols, C. H. Peck, D. Kong, R. Miller, and C. Long. Interacting at a distance: measuring the performance of laser pointers and other devices. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '02)*, pp. 33–40. ACM Press, New York, NY, USA: 2002.

[142] J.-L. Nespoulous, P. Perron, and A. R. Lecours. *The biological foundations of gesture: motor and semiotic aspects*. Neuropsychology and Neurolinguistics. Lawrence Erlbaum Associates, Hillsdale, N.J.: 1986.

[143] A. Neto and C. Duarte. Comparing Gestures and Traditional Interaction Modalities on Large Displays. In: *Human-Computer Interaction (INTERACT '09)*, vol. 5727/2009 of *Lecture Notes in Computer Science*, pp. 58–61. Springer Berlin / Heidelberg: 2009.

[144] K. Nickel and R. Stiefelhagen. Visual recognition of pointing gestures for human-robot interaction. *Image and Vision Computing*, **vol. 25 (12)**: 1875–1884: 2007.

[145] J. Nielsen. A virtual protocol model for computer-human interaction. *International journal of man-machine studies*, **vol. 24 (3)**: 301–312: 1986.

[146] M. Nielsen, M. Störring, T. Moeslund, and E. Granum. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In: *Gesture-Based Communication in Human-Computer Interaction*, vol. 2915/2004 of *Lecture Notes in Computer Science*, pp. 409–420. Springer Berlin / Heidelberg: 2004.

[147] A. Nijholt. Multimodality and Ambient Intelligence. In: W. Verhaegh, E. Aarts, and J. Korst, eds., *Algorithms in Ambient Intelligence*, pp. 21–53. Kluwer Academic Publishers, Boston/Dordrecht/London: 2003.

[148] A. Nijholt, R. op den Akker, and D. Heylen. Meetings and meeting modeling in smart environments. *AI and Society*, **vol. 20 (2)**: 202–220: 2006.

[149] A. Nijholt, D. Reidsma, and R. Poppe. Games and Entertainment in Ambient Intelligence Environments. In: H. Aghajan, R. Delgado, and J. C. Augusto, eds., *Human-Centric Interfaces for Ambient Intelligence*. Elsevier: 2009.

[150] A. Nijholt, D. Reidsma, Z. Ruttkay, H. van Welbergen, and P. Bos. Nonverbal and Bodily Interaction in Ambient Entertainment. In: *Proceedings workshop on Fundamentals of Verbal and Non-verbal Communication and the Biometrical Issue, Vietri sul Mare, Italy*, vol. 18 of *NATO Security through Science Series, E: Human and Societal Dyna*, pp. 343–348. IOS Press, Amsterdam: 2007.

[151] M. Nixon, B. McCallum, R. Fright, and B. Price. The Effects of Metals and Interfering Fields on Electromagnetic Trackers. *Presence: Teleoperators and Virtual Environments*, **vol. 7 (2)**: 204–218: 1998.

[152] D. Norman. *The Design of Everyday Things*. Doubleday, New York: 1988.

[153] D. Norman. Cognitive artifacts. In: J. Carroll, ed., *Designing Interaction: Psychology at the Human-Computer Interface*, pp. 17–38. Cambridge University Press, Cambridge, UK: 1991.

[154] D. Norman and S. Draper. *User Centered System Design; New Perspectives on Human-Computer Interaction*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, USA: 1986.

[155] S. Ong and S. Ranganath. Automatic Sign Language Analysis: A Survey and the Future

beyond Lexical Meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **vol. 27 (6)**: 873–891: 2005.

[156] R. Ott, D. Thalmann, and F. Vexo. Haptic feedback in Mixed-Reality Environments. *The Visual Computer: International Journal of Computer Graphic*, **vol. 23 (9)**: 843–849: 2007.

[157] S. Oviatt. Toward Adaptive Information Fusion in Multimodal Systems. In: *Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop (MMUI '05)*, vol. 57 of *ACM International Conference Proceeding Series*, pp. 15–27: 2006.

[158] S. Oviatt and P. Cohen. Perceptual user interfaces: multimodal interfaces that process what comes naturally. *Communications of the ACM*, **vol. 43 (3)**: 45–53: 2000.

[159] R. Owen, G. Kurtenbach, G. Fitzmaurice, T. Baudel, and W. Buxton. When it gets more difficult, use both hands: exploring bimanual curve manipulation. In: *Proceedings of the 2005 conference on Graphics interface (GI '05)*, vol. 112, pp. 17–24. Canadian Human-Computer Communications Society, University of Waterloo, Waterloo, Ontario, Canada: 2005.

[160] M. Pantic, A. Pentland, A. Nijholt, and T. Huang. Human Computing and Machine Understanding of Human Behavior: A Survey. In: *Proceedings of the 8th international conference on Multimodal interfaces (ICMI '06)*, vol. 8, pp. 239–248. ACM, ACM Press: 2006.

[161] A.-Y. Park and S.-W. Lee. Gesture Spotting in Continuous Whole Body Action Sequences Using Discrete Hidden Markov Models. In: [58], chap. 12, pp. 100–111.

[162] E. Park, B. Kim, W. Salim, and A. Cheok. Magic Asian art. In: *Extended abstracts on Human factors in computing systems (CHI '06)*, pp. 255–258. ACM, New York, NY, USA: 2006.

[163] J. Park and Y.-L. Yoon. LED-Glove Based Interactions in Multi-Modal Displays for Teleconferencing. In: *16th International Conference on Artificial Reality and Telexistence*, pp. 395–399: 2006.

[164] H. Patel, O. Stefani, S. Sharples, H. Hoffmann, I. Karaseitanidis, and A. Amditis. Human centred design of 3-D interaction devices to control virtual environments. *International Journal of Human-Computer Studies*, **vol. 64 (3)**: 207 – 220: 2006.

[165] J. Patten, B. Recht, and H. Ishii. Interaction Techniques for Musical Performance with Tabletop Tangible Interfaces. *Advances in Computer Entertainment Technology*: 2006.

[166] V. Pavlovic, R. Sharma, and T. Huang. Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **vol. 19 (7)**: 677–695: 1997.

[167] T. Pfeiffer, M. E. Latoschik, and I. Wachsmuth. Conversational Pointing Gestures for Virtual Reality Interaction: Implications from an Empirical Study. In: *Virtual Reality Conference*, pp. 281–282: 2008.

[168] R. Poppe. *Discriminative Vision-Based Recovery and Recognition of Human Motion*. Ph.D. thesis, University of Twente: 2009.

[169] T. Prante, C. Rocker, N. Streitz, R. Stenzel, C. Magerkurth, D. van Alphen, and D. Plewe. Hello. Wall - Beyond Ambient Displays. In: *Video Track and Adjunct Proceedings of the 5th Intern. Conference on Ubiquitous Computing (UBICOMP'03)*, pp. 1–2: 2003.

[170] S. Prillwitz, R. Leven, H. Zienert, T. Hanke, and J. Henning. *Hamburg Notation System for Sign Languages - An introductory guide*, vol. 5. Signum, Hamburg, Germany: 1989.

[171] T. Pylvänäinen. Accelerometer Based Gesture Recognition Using Continuous HMMs. In: *Pattern Recognition and Image Analysis*, vol. 3522/2005 of *Lecture Notes in Computer Science*, pp. 639–646. Springer: 2005.

[172] F. Quek. Unencumbered gestural interaction. *IEEE Multimedia*, **vol. 3 (4)**: 36–47: 1996.

[173] F. Quek, D. McNeill, R. Ansari, X.-F. Ma, R. Bryll, S. Dunn, and K.-E. McCullough. Gesture cues for conversational interaction in monocular video. In: *Proceedings of the International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pp. 119–126. IEEE Computer Society: 1999.

[174] F. Quek, D. McNeill, R. Bryll, S. Duncan, X.-F. Ma, C. Kirbas, K.-E. McCullough, and R. Ansari. Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction*, **vol. 9 (3)**: 171–193: 2002.

[175] F. Rauscher, R. Krauss, and Y. Chen. Gesture, Speech, and Lexical Access: The Role of Lexical Movements in Speech Production. *Psychological Science*, **vol. 7 4**: 226–231: 1996.

[176] H. Rauwerda, W. de Leeuw, J. Adriaanse, M. Bouwhuis, P. van der Vet, and T. Breit. The Role of e-BioLabs in a Life Sciences Collaborative Working Environment. In: *Proceedings of the eChallenges 2007*: 2007.

[177] H. Rauwerda, M. Roos, B. Hertzberger, and T. Breit. The promise of a virtual lab in drug discovery. *Drug Discovery Today*, **vol. 11**: 228–236: 2006.

[178] H. Rauwerda, P. van der Vet, O. Kulyk, I. Wassink, W. Fikkert, W. de Leeuw, J. Adriaanse, B. van Dijk, G. van der Veer, M. Bouwhuis, T. Breit, and A. Nijholt. E-science support for omics experimentation. In: *Benelux Bioinformatics Conference*: 2007.

[179] B. Reeves and C. Nash. *The media equation: How people treat computers, television, and new media like real people and places*. CSLI Publications, Cambridge University Press: 1996.

[180] W. Reisig. *Petri nets: an introduction*. Springer-Verlag New York, Inc., New York, NY, USA: 1985.

[181] J. Rekimoto. SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '02)*, pp. 113–120. ACM Press, New York, NY, USA: 2002.

[182] I. Rock, D. Wheeler, and L. Tudor. Can we imagine how objects look from other viewpoints? *Cognitive Psychology*, **vol. 21 (2)**: 185 – 210: 1989.

[183] S. Rusdorf and G. Brunnett. Real time tracking of high speed movements in the context of a table tennis application. In: *Proceedings of the ACM symposium on Virtual reality software and technology (VRST '05)*, pp. 192–200. ACM Press: 2005.

[184] H. Sagawa and M. Takeuchi. A Method for Recognizing a Sequence of Sign Language Words Represented in a Japanese Sign Language Sentence. In: *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000 (FG '00)*, p. 434. IEEE Computer Society, Washington, DC, USA: 2000.

[185] M. Schena. *Microarray Analysis*. John Wiley and Sons Inc., Hoboken, New Jersey, 1 ed.: 2003.

[186] S. Schkolne, M. Pruett, and P. Schröder. Surface drawing: creating organic 3D shapes with the hand and tangible tools. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '01)*, pp. 261–268. ACM Press, New York, NY, USA: 2001.

[187] M. Schlattman and R. Klein. Simultaneous 4 gestures 6 DOF real-time two-hand tracking without any markers. In: *Proceedings of the 2007 ACM symposium on Virtual reality software and technology (VRST '07)*, pp. 39–42. ACM, New York, NY, USA: 2007.

[188] J. Schöning, F. Daiber, and A. Krüger. Advanced Navigation Techniques for Spatial Information Using Whole Body Motion. In: *Workshop on Whole Body Interation: The Future of the Human Body (HCI '08)*, p. 2: 2008.

[189] J. Schöning and A. Krüger. Multi-Modal Navigation through Spatial Information. In: *Proceedings of the 5th International Conference on Geographic Information Science (GIScience '08)*, pp. 151–154: 2008. Extended Abstracts.

[190] B. Shneiderman. Direct manipulation for comprehensible, predictable and controllable user interfaces. In: *Proceedings of the 2nd international conference on Intelligent user interfaces (IUI '97)*, pp. 33–39. ACM Press, New York, NY, USA: 1997.

[191] B. Shneiderman and P. Maes. Direct manipulation vs. interface agents. *interactions*, **vol. 4 (6)**: 42–61: 1997.

[192] N. Sisson. Dialogue management reference model. *SIGCHI Bulletin*, **vol. 18 (2)**: 34–35: 1986.

[193] J.-J. Song, S. K. Smith, G. J. Hannon, and L. Joshua-Tor. Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity. *Science*, **vol. 305 (5689)**: 1434–1437: 2004.

[194] T. Sowa. The Recognition and Comprehension of Hand Gestures - A Review and Research Agenda. In: I. Wachsmuth and G. Knoblich, eds., *Modeling Communication with Robots and Virtual Humans*, vol. 4930/2008 of *Lecture Notes in Computer Science*, chap. 3, pp. 38–56. Springer Berlin / Heidelberg: 2008.

[195] T. Sowa and I. Wachsmuth. Interpretation of Shape-Related Iconic Gestures in Virtual Environments. In: *Gesture and Sign Language in Human-Computer Interaction*, vol. 2298/2002 of *Lecture Notes in Computer Science*, chap. 3, pp. 51–74. Springer Berlin / Heidelberg: 2002.

[196] O. Stefani and J. Rauschenbach. 3D input devices and interaction concepts for optical tracking in immersive environments. In: *Proceedings of the workshop on Virtual environments 2003 (EGVE '03)*, pp. 317–318. ACM, New York, NY, USA: 2003.

[197] M. Streit. Why Are Multimodal Systems so Difficult to Build? - About the Difference between Deictic Gestures and Direct Manipulation. In: *Proceedings of the 2nd International Conference on Cooperative Multimodal Communication (CMC '98)*, pp. 176–196: 1998.

[198] N. Streitz and P. Nixon. The dissappearing computer. *Communications of the ACM*, **vol. 48 (3)**: 32–35: 2005.

[199] J. A. Sturm. *On the Usability of Multimodal Interaction for Mobile Access to Information Services*. Ph.D. thesis, Raboud University, Nijmegen, The Netherlands: 2005.

[200] K. Swaminathan and S. Sato. Interaction design for large displays. *interactions*, **vol. 4 (1)**: 15–24: 1997.

[201] J. R. Tame and B. Vallone. The structures of deoxy human haemoglobin and the mutant Hb Tyralpha42His at 120 K. *Acta crystallographica. Section D, Biological crystallography*, **vol. 56 (Pt 7)**: 805–811: 2000.

[202] G. ten Holt, J. Arendsen, H. de Ridder, A. Koenderink-van Doorn, M. Reinders, and E. Hendriks. Sign language perception research for improving automatic sign language recognition. In: B. Rogowitz and T. Pappas, eds., *Proceedings of the SPIE (Human Vision and Electronic Imaging XIV)*, vol. 7240, p. 72400C. SPIE: 2009.

[203] B. Tognazzini. The "Starfire" video prototype project: a case history. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '94)*, pp. 99–105. ACM, New York, NY, USA: 1994.

[204] E. Tse, S. Greenberg, C. Shen, and C. Forlines. Multimodal Multiplayer Tabletop Gaming. *Computers in Entertainment (CIE)*, **vol. 5 (2)**: 12: 2006.

[205] E. Tse, C. Shen, S. Greenberg, and C. Forlines. Enabling interaction with single user appli-

cations through speech and gestures on a multi-user tabletop. In: *Proceedings of the working conference on Advanced visual interfaces (AVI '06)*, pp. 336–343. ACM Press, New York, NY, USA: 2006.

[206] E. Tse, C. Shen, S. Greenberg, and C. Forlines. How pairs interact over a multimodal digital table. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '07)*, pp. 215–218. ACM, New York, NY, USA: 2007.

[207] A. van Dam. Post-WIMP user interfaces. *Communications of the ACM*, **vol. 40 (2)**: 63–67: 1997.

[208] F. van der Sluis. *Multimodal Reference, Studies in Automatic Generation of Multimodal Referring Expressions*. Ph.D. thesis, University of Tilburg, The Netherlands: 2005.

[209] P. van der Vet, O. Kulyk, I. Wassink, W. Fikkert, H. Rauwerda, B. van Dijk, G. van der Veer, T. Breit, and A. Nijholt. Smart Environments for Collaborative Design, Implementation, and Interpretation of Scientific Experiments. In: *Workshop on AI for Human Computing (AI4HC)*, vol. 20 of *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 79–86. AAAI Press: 2007.

[210] J. van Gelder. Afstandsbediening wordt mediahype. *Cursor*, **vol. 50 (27)**: 6–7: 2008.

[211] B. van Schooten, B. van Dijk, E. Zudilova-Seinstra, P. de Koning, and H. Reiber. Evaluating visualisation and navigation techniques for interpretation of MRA data. In: *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP '09)*, pp. 405–408. INSTICC Press, Portugal: 2009.

[212] H. van Welbergen and Z. Ruttkay. On the Parametrization of Clapping. In: *Gesture-Based Human-Computer Interaction and Simulation*, chap. 4, pp. 36–47. Springer Berlin / Heidelberg: 2009.

[213] A. Vasilakos and W. Pedrycz. *Ambient Intelligence, Wireless Networking, And Ubiquitous Computing*. Artech House, Inc., Norwood, MA, USA: 2006.

[214] D. Vogel and R. Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In: *Proceedings of the 17th annual ACM symposium on User interface software and technology (UIST '04)*, pp. 137–146. ACM Press, New York, NY, USA: 2004.

[215] D. Vogel and R. Balakrishnan. Distant freehand pointing and clicking on very large, high resolution displays. In: *Proceedings of the 18th annual ACM symposium on User interface software and technology (UIST '05)*, pp. 33–42. ACM Press, New York, NY, USA: 2005.

[216] J. Wachs, H. Stern, Y. Edan, M. Gillam, C. Feied, M. Smith, and J. Handler. Gestix: A Doctor-Computer Sterile Gesture Interface for Dynamic Environments. In: *Soft Computing in Industrial Applications*, vol. 39/2007 of *Advances in Soft Computing*, pp. 30–39. Springer Berlin / Heidelberg: 2007.

[217] J. Wachs, H. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, and M. Smith. A Gesture-based Tool for Sterile Browsing of Radiology Images. *Journal of the American Medical Informatics Association*, **vol. 15 (3)**: 321 – 323: 2008.

[218] I. Wachsmuth and M. Fröhlich, eds. *Gesture and Sign Language in Human-Computer Interaction*, vol. 1371/1997 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg: 1997.

[219] A. Waern, M. Montola, and J. Stenros. The three-sixty illusion: designing for immersion in pervasive games. In: *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*, pp. 1549–1558. ACM, New York, NY, USA: 2009.

[220] C. Wagner, S. J. Lederman, and R. Howe. A tactile shape display using RC servomotors. In: *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS '02)*, pp. 354–355: 2002.

[221] W. Wahlster. *SmartKom: Foundations of Multimodal Dialogue Systems*, vol. XVIII of *Cognitive Technologies*. Springer: 2006.

[222] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, **vol. 36**: 585–601: 2003.

[223] Y. Wang and C. MacKenzie. The role of contextual haptic and visual constraints on object manipulation in virtual environments. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '00)*, pp. 532–539. ACM, New York, NY, USA: 2000.

[224] G. Welch and G. Bishop. An Introduction to the Kalman Filter. *Technical Report TR95-041*, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA: 1995.

[225] P. Wellner. Interacting with paper on the DigitalDesk. *Communications of the ACM*, **vol. 36 (7)**: 87–96: 1993.

[226] A. Wexelblat. An approach to natural gesture in virtual environments. *ACM Transactions on Computer-Human Interaction*, **vol. 2 (3)**: 179–200: 1995.

[227] A. Wexelblat. Research Challenges in Gesture: Open Issues and Unsolved Problems. In: *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*, vol. 1371, pp. 1–11. Springer-Verlag, London, UK: 1998.

[228] A. Wilson. TouchLight: an imaging touch screen and display for gesture-based interaction. In: *Proceedings of the 6th international conference on Multimodal interfaces (ICMI '04)*, pp. 69–76. ACM Press, New York, NY, USA: 2004.

[229] A. Wilson. Robust computer vision-based detection of pinching for one and two-handed gesture input. In: *Proceedings of the 19th annual ACM symposium on User interface software and technology (UIST '06)*, pp. 255–258. ACM, New York, NY, USA: 2006.

[230] J. Wobbrock, M. Morris, and A. Wilson. User-defined gestures for surface computing. In: *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*, pp. 1083–1092. ACM, New York, NY, USA: 2009.

[231] J. Wobbrock, B. Myers, and H. H. Aung. The performance of hand postures in front- and back-of-device interaction for mobile computing. *International Journal of Human-Computer Studies*, **vol. 66 (12)**: 857–875: 2008.

[232] J. Wobbrock, J. Rubinstein, M. Sawyer, and A. Duchowski. Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In: *Proceedings of the 2008 symposium on Eye tracking research & applications (ETRA '08)*, pp. 11–18. ACM, New York, NY, USA: 2008.

[233] D. Wolpert and Z. Ghahramani. Computational principles of movement neuroscience. *Nature Neuroscience*, **vol. 3**: 1212–1217: 2000.

[234] D. Wormell and E. Foxlin. Advancements in 3D interactive devices for virtual environments. In: *Proceedings of the workshop on Virtual environments (EGVE '03)*, pp. 47–56. ACM, New York, NY, USA: 2003.

[235] C. Wu, H. Aghajan, and R. Kleihorst. Real-Time Human Posture Reconstruction in Wireless Smart Camera Networks. In: *Proceedings of the 7th international conference on Information processing in sensor networks (IPSN '08)*, pp. 321–331. IEEE Computer Society, Washington, DC, USA: 2008.

[236] M. Wu and R. Balakrishnan. Multi-finger and whole hand gestural interaction techniques

for multi-user tabletop displays. In: *Proceedings of the 16th annual ACM symposium on User interface software and technology (UIST '03)*, pp. 193–202. ACM Press, New York, NY, USA: 2003.

[237] M. Wu, C. Shen, K. Ryall, C. Forlines, and R. Balakrishnan. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In: *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TableTop '06)*, pp. 185–192: 2006.

[238] Y. Wu and T. Huang. Hand modeling, analysis and recognition. *IEEE Signal Processing Magazine*, **vol. 18 (3)**: 51–60: 2001.

[239] I. Yoda, K. Sakaue, and T. Inoue. Development of head gesture interface for electric wheelchair. In: *Proceedings of the 1st international convention on Rehabilitation engineering & assistive technology (i-CREATe '07)*, pp. 77–80. ACM, New York, NY, USA: 2007.

[240] Y.-R. Yuan, Y. Pei, H.-Y. Chen, T. Tuschl, and D. Patel. A Potential Protein-RNA Recognition Event along the RISC-Loading Pathway from the Structure of A. aeolicus Argonaute with Externally Bound siRNA. *Structure*, **vol. 14 (10)**: 1557 – 1565: 2006.

[241] X. Zhang and S. MacKenzie. Evaluating Eye Tracking with ISO 9241 - Part 9. In: *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, vol. 4552/2007 of *Lecture Notes in Computer Science*, chap. 85, pp. 779–788. Springer Berlin / Heidelberg: 2007.

[242] T. Zimmerman, J. Smith, J. Paradiso, D. Allport, and N. Gershenfeld. Applying electric field sensing to human-computer interfaces. In: *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '95)*, pp. 280–287. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA: 1995.

Appendices

Appendix A

Gestures Descriptions

This appendix contains the descriptions that explained the gestures to the participants in our large-scale online investigation and its two validation conditions.

Pointing

***Ray-casting*:** The cursor is precisely at the point where you are pointing at with your index finger.

***Repetitive taps*:** You make repetitive pointing taps with your index finger to move the cursor in the direction where your hand is oriented. It is similar to using a keyboard's arrow keys repetitively.

***Tap once*:** You point upwards with your index finger to let the cursor move upwards on the screen. Moving in other directions is identical. The technique is just like holding a keyboard arrow key.

Selecting

***AirTap*:** Tapping with the index finger just like clicking the left mouse button.

***ThumbTrigger*:** While pointing at a target with the index finger, using the thumb to tap against the middle finger.

***Dwelling*:** Keeping the index finger pointing still on a target of your choice for some time, for example, 300 milliseconds.

***Encircling*:** Using the index finger to point you make a circular shape around the target that you wish to select.

***FistGrab*:** While pointing at a target with the index finger you close your hand to select.

Deselecting

***DropIt*:** While having a target selection you open your hand with the palm down as if dropping the target on the floor.

***Retract to rest*:** While having a target selected you retract your hand to rest besides your body.

***Jerky retract*:** While having a target selected, you briefly jerk your hand backwards.

***Select other*:** You deselect by selecting another target, for example, by tapping on it, or by selecting no target at all (an empty area on the display).

Resizing: enlarging and shrinking

***Fingers apart*:** You enlarge a target window/picture by moving two fingers of one hand apart. To shrink a target you perform the opposite: moving two fingers of one hand toward each other.

***Hands apart*:** You enlarge a target window/picture by moving your two hands apart. To shrink a target you perform the opposite: moving your two hands toward each other.

***PullPush*:** You close your hand to grab the display and by pulling it closer you enlarge it, pushing it away will shrink it.

***Referenced PullPush*:** You 'grab' the display with the palm towards yourself. By moving your second hand relative to the first one, you resize the target.

Activate and Deactivate

***AirTap & exit cross*:** You select a target by tapping on it with your index finger, just as with the left mouse button. A second tap activates the target. A tap on the cross in the top of the window deactivates it.

***AirTap*:** You select a target by tapping on it with your index finger, just as with the left mouse button. A second tap activates the target, a third deactivates it.

***ThumbTrigger*:** You select a target by tapping your thumb on your middle finger while pointing with the index finger. A second tap activates the target, a third deactivates it.

***Dwell & exit cross*:** You activate a target by pointing at it for a short amount of time. When activated, you deactivate the target by briefly pointing at the cross in the top of the window.

***Jerky PullPush*:** You activate the target by pulling it towards yourself in a short jerky fashion. Deactivation is the opposite: you push it away in a short jerky fashion.

***Open palm facing*:** You activate the target by showing a flat hand with the palm towards face. To deactivate, you turn the flat hand with the palm towards to display.

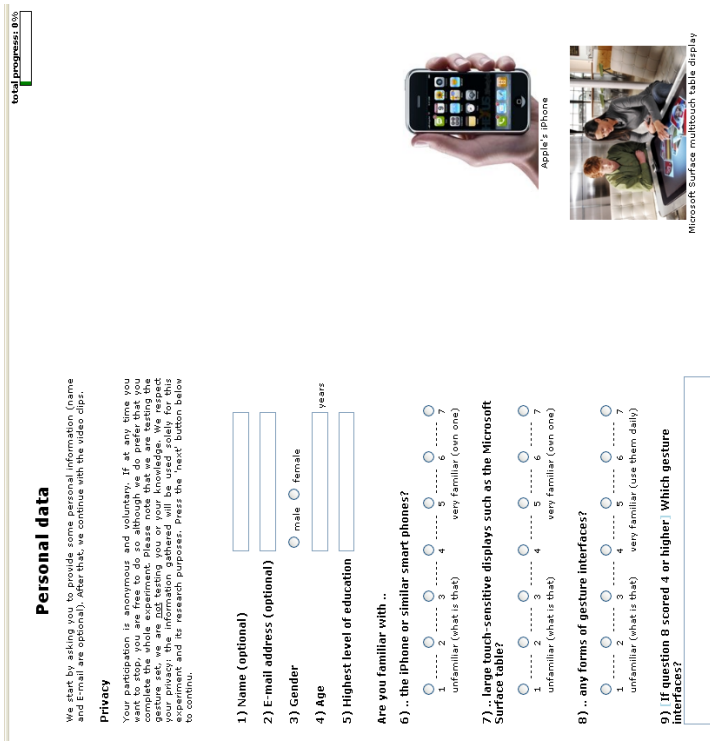
***Activation and deactivation zones*:** You activate a target by dragging it to an activation zone on the display. To deactivate the target you drag it to a deactivation zone on the display in much the same manner.

***Drawing 'play' and 'stop' shapes*:** You activate the target that you are pointing at by drawing a triangle shape, similar in shape to the play button on a DVD player. To deactivate the target, you draw a rectangle shape, similar to the stop button on a DVD player.

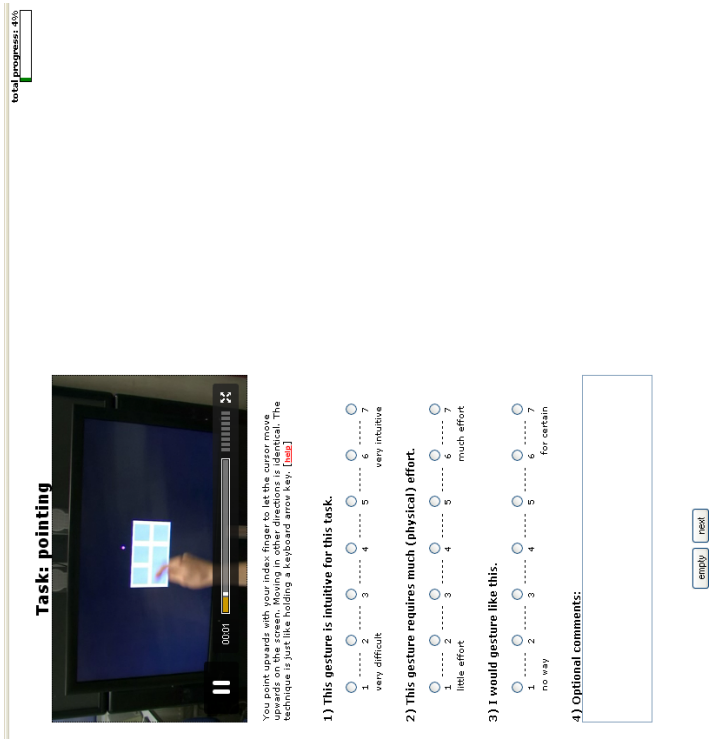
Context Menu

***Clapping*:** By clapping your hands you activate a option menu within the current context. Clapping a second time will close this menu again.

***PinkieTrigger*:** You open an options menu. Tap your thumb on your pinky finger. This is comparable to pressing the right mouse button. When you tap your thumb on your pinky finger a second time the menu closes again.



(a)



(b)

Figure A.1: Screenshots of the online questionnaire at `http://fikkert.net`: (a) the personal data that we gathered from our participants. (b) the illustrated gestures.

Appendix B

Prototype

This appendix contains the questionnaire that was filled out by the participants of the prototype evaluation that was described in Chapter 7. In addition, it displays the electrical schema that was implemented in the two gloves.

B.1 Questionnaire - part 1

These forms were filled out before beginning the experiment.

For experimenter only
Participant: ---------
Start/end: -------

Questionnaire – Personal information
Thank you for participating in this investigation. We will start by asking you about some personal information. Please fill out the following questions.

Q1. Gender: O male
O female

Q2. Handedness: O right handed
O left handed
O ambidextrous

Q3. Age: ----------

Q4. Education: ----------

Q5. Average hours at PC per day: ----------

Q6. Participated in: O Wizard of Oz – gesturing to control a map of Twente
O Online questionnaire – evaluating gestures in video clips
O Validation – LED on your finger for pointing

Q7. Know molecule: O ice
(see color sheet) O 1c3n
O 1u04
O 2f8s

Q8. Experience with pen-based devices such as smart phones, PDAs, tablet PCs, etc.

O---------O---------O---------O---------O---------O---------O
1 2 3 4 5 6 7
never used one dialy use

Q9. (if Q8 scored 4 or higher) Which ones?

Q10. Experience with the Apple iPhone

O---------O---------O---------O---------O---------O---------O
1 2 3 4 5 6 7
never used one dialy use

Q11. Experience with other multi-touch systems

O---------O---------O---------O---------O---------O---------O
1 2 3 4 5 6 7
never used one dialy use

Figure B.1: The questionnaire (page 1) used to evaluate the prototype.

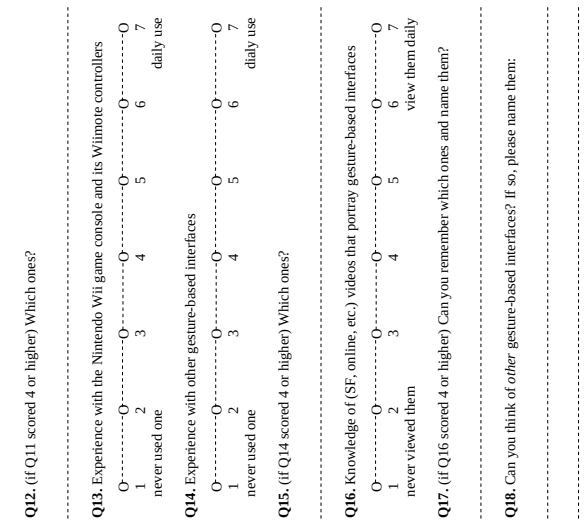


Figure B.2: The questionnaire (page 2) used to evaluate the prototype.

B.2 Questionnaire - part 2

These forms were filled out after completing the whole experiment. They address the gesture interaction as a whole.

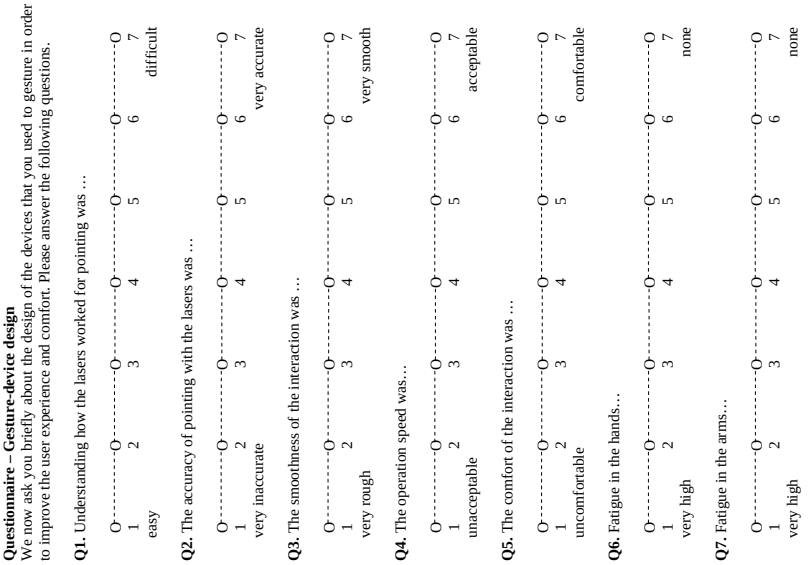


Figure B.3: The questionnaire (page 3) used to evaluate the prototype.

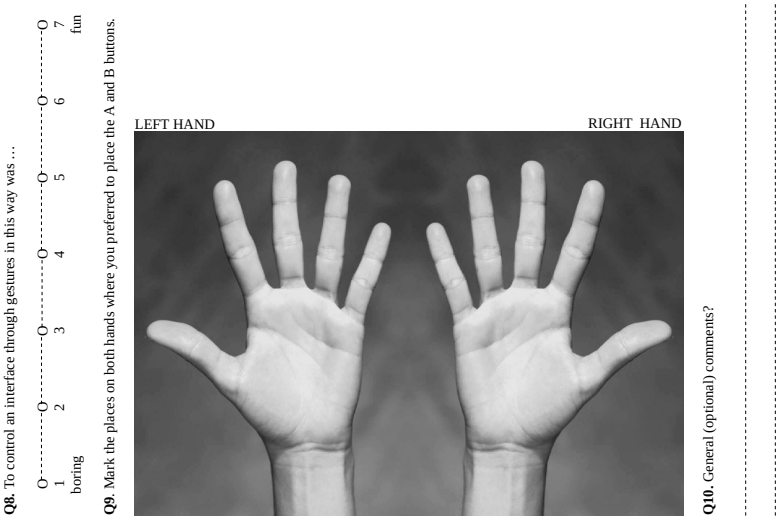


Figure B.4: The questionnaire (page 4) used to evaluate the prototype.

B.3 Questionnaire - part 3

These forms were filled out after completing the whole experiment. They address the commands that could be issued separately.

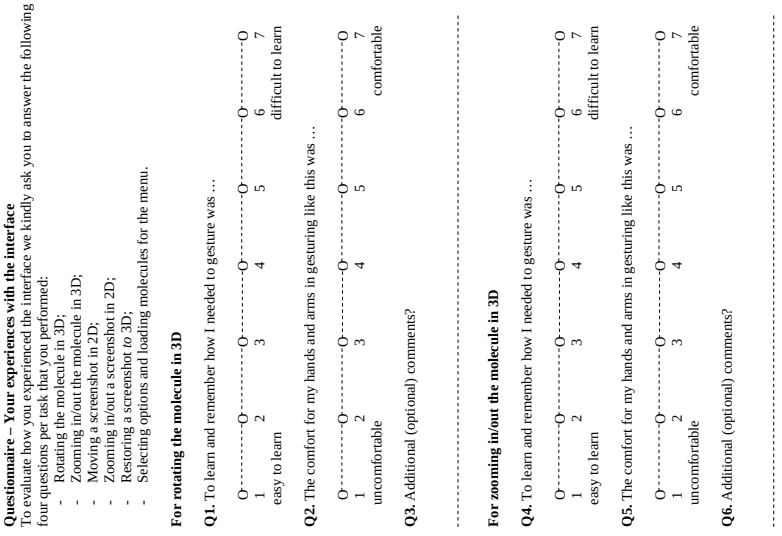


Figure B.5: The questionnaire (page 5) used to evaluate the prototype.

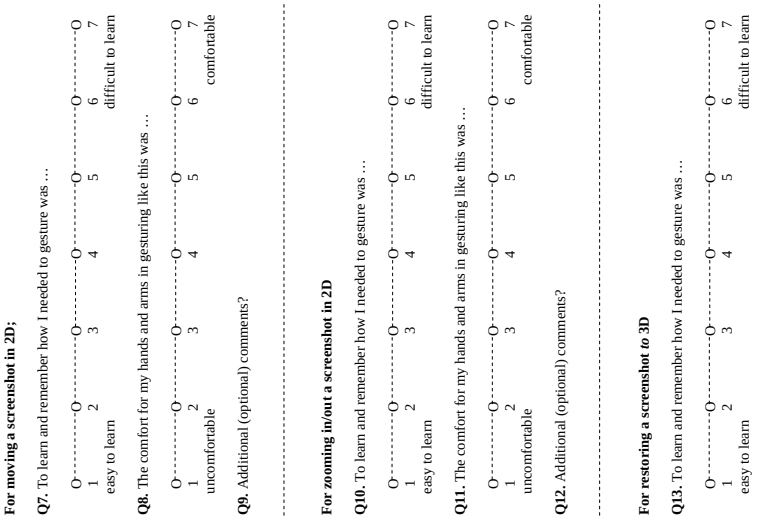


Figure B.6: The questionnaire (page 6) used to evaluate the prototype.

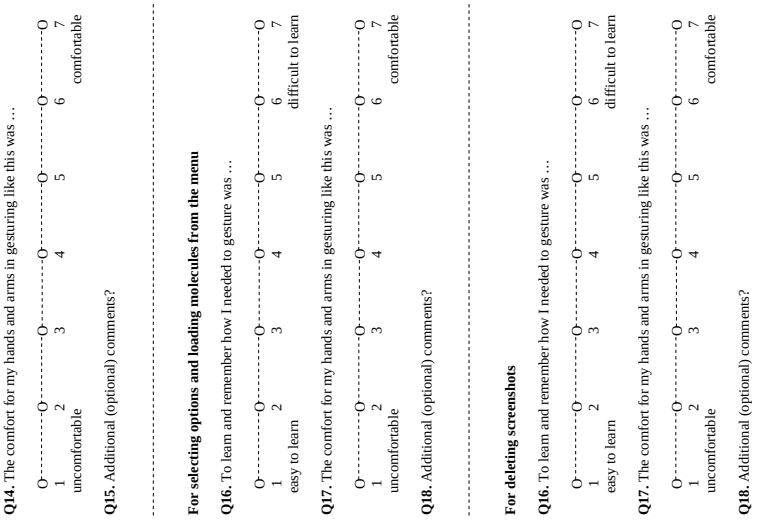


Figure B.7: The questionnaire (page 7) used to evaluate the prototype.

B.4 Questionnaire results

	mean	std. dev.	variance	kurtosis	skewness	K^2	p
avg. hours at PC	7.9	2.3	5.2	-0.4	-0.7	2.3	.31
pen-based devices	4.3	2.2	4.9	-1.8	0.2	3.0	.22
iPhone	2.9	2.3	5.5	-0.4	1.1	5.6	.06
other multi-touch	3.4	1.8	3.1	-0.4	0.7	2.6	.27
Wii(mote)	3.8	1.6	2.6	-0.8	0.3	1.4	.49
other gesture int.	2.5	1.7	2.9	1.4	1.5	9.9	< .01
video clips	3.5	1.5	2.3	-1.1	0.3	1.9	.39

Table B.1: Experience of our subjects before taking part in the investigation (N = 23).

	mean	std. dev.	variance	kurtosis	skewness	K^2	p
how lasers worked	6.4	0.8	0.6	2.9	-1.6	13.7	< .01
pointing accuracy	5.0	1.1	1.1	1.9	-1.3	9.6	< .01
interaction smoothness	3.9	1.2	1.5	-0.3	0.1	0.4	.82
operation speed	4.6	1.2	1.4	-0.1	-0.2	0.3	.86
interaction comfort	5.1	1.1	1.1	3.0	-1.3	11.3	< .01
fatigue in hands	6.0	1.3	1.7	2.7	-1.6	12.9	< .01
fatigue in arms	5.6	1.2	1.5	2.6	-1.4	11.5	< .01
fun or boring?	6.1	1.0	1.1	1.9	-1.3	9.0	.01

Table B.2: Overall interaction ratings of the experience during the experiment: see Appendix B.2 for the complete questions. Scoring was adjusted so that 1: negative score (worst), 7: positive score (best) (N = 23).

command	question	mean	std. dev.	variance	kurtosis	skewness	K^2	p
rotate 3D	learn gestures	5.9	1.4	1.8	2.2	-1.6	11.8	< .01
	gesture comfort	5.7	1.2	1.4	3.8	-1.9	17.9	< .01
zoom 3D	learn gestures	6.0	0.8	0.7	0.2	-0.6	1.9	.38
	gesture comfort	5.7	1.0	1.1	1.0	-1.1	6.4	.04
move 2D	learn gestures	6.7	0.5	0.2	-1.29	-0.9	5.4	0.1
	gesture comfort	6.0	0.8	0.6	2.2	-1.2	8.9	0.7
zoom 2D	learn gestures	6.4	0.6	0.3	-0.7	-0.3	1.2	.55
	gesture comfort	5.9	1.0	1.0	2.4	-1.3	10.5	< .01
restore	learn gestures	5.4	1.3	1.6	-0.7	-0.5	2.1	.35
	gesture comfort	5.7	1.1	1.2	0.6	-1.0	4.9	.09
options	learn gestures	6.8	0.4	0.2	0.2	-1.5	7.9	.01
	gesture comfort	6.3	0.9	0.7	2.5	-1.5	12.0	< .01
delete	learn gestures	5.5	1.2	1.4	-0.8	-0.3	1.5	.46
	gesture comfort	5.9	0.9	0.8	0.9	-1.1	5.7	.06

Table B.3: Detailed interaction ratings, per command that could be given. The commands are described in Section 7.1.3. Scoring was adjusted so that 1: negative score (worst), 7: positive score (best). (N = 23).

SIKS dissertation series

Since 1998, all dissertations written by Ph.D. students who have conducted their research under auspices of a senior research fellow of the SIKS research school are published in the SIKS Dissertation Series. This thesis is the 251st in the series.

2010-07 Wim Fikkert (UT), *Gesture Interaction at a Distance*
2010-06 Sander Bakkes (UvT), *Rapid Adaptation of Video Game AI*
2010-05 Claudia Hauff (UT), *Predicting the Effectiveness of Queries and Retrieval Systems*
2010-04 Olga Kulyk (UT), *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments*
2010-03 Joost Geurts (CWI), *A Document Engineering Model and Processing Framework for Multimedia documents*
2010-02 Ingo Wassink (UT), *Work flows in Life Science*
2010-01 Matthijs van Leeuwen (UU), *Patterns that Matter*
2009-46 Loredana Afanasiev (UvA), *Querying XML: Benchmarks and Recursion*
2009-45 Jilles Vreeken (UU), *Making Pattern Mining Useful*
2009-44 Roberto Santana Tapia (UT), *Assessing Business-IT Alignment in Networked Organizations*
2009-43 Virginia Nunes Leal Franqueira (UT), *Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients*
2009-42 Toine Bogers (UvT), *Recommender Systems for Social Bookmarking*
2009-41 Igor Berezhnyy (UvT), *Digital Analysis of Paintings*
2009-40 Stephan Raaijmakers (UvT), *Multinomial Language Learning: Investigations into the Geometry of Language*
2009-39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin), *Service Substitution – A Behavioral Approach Based on Petri Nets*
2009-38 Riina Vuorikari (OU), *Tags and self-organisation: a metadata ecology for learning resources in a multilingual context*
2009-37 Hendrik Drachsler (OUN), *Navigation Support for Learners in Informal Learning Networks*
2009-36 Marco Kalz (OUN), *Placement Support for Learners in Learning Networks*
2009-35 Wouter Koelewijn (UL), *Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling*
2009-34 Inge van de Weerd (UU), *Advancing in Software Product Management: An Incremental Method Engineering Approach*
2009-33 Khiet Truong (UT), *How Does Real Affect Affect Affect Recognition In Speech?*
2009-32 Rik Farenhorst (VU) and Remco de Boer (VU), *Architectural Knowledge Management: Supporting Architects and Auditors*
2009-31 Sofiya Katrenko (UVA), *A Closer Look at Learning Relations from Text*
2009-30 Marcin Zukowski (CWI), *Balancing vectorized query execution with bandwidth-optimized storage*
2009-29 Stanislav Pokraev (UT), *Model-Driven Semantic Integration of Service-Oriented Applications*
2009-27 Christian Glahn (OU), *Contextual Support of social Engagement and Reflection on the Web*
2009-26 Fernando Koch (UU), *An Agent-Based Model for the Development of Intelligent Mobile Services*
2009-25 Alex van Ballegooij (CWI), *RAM: Array Database Management through Relational Mapping*
2009-24 Annerieke Heuvelink (VUA), *Cognitive Models for Training Simulations*
2009-23 Peter Hofgesang (VU), *Modelling Web Usage in a Changing Environment*
2009-22 Pavel Serdyukov (UT), *Search For Expertise: Going beyond direct evidence*
2009-21 Stijn Vanderlooy (UM), *Ranking and Reliable Classification*
2009-20 Bob van der Vecht (UU), *Adjustable Autonomy: Controling Influences on Decision Making*
2009-19 Valentin Robu (CWI), *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets*
2009-18 Fabian Groffen (CWI), *Armada, An Evolving Database System*
2009-17 Laurens van der Maaten (UvT), *Feature Extraction from Visual Data*
2009-16 Fritz Reul (UvT), *New Architectures in Computer Chess*
2009-15 Rinke Hoekstra (UVA), *Ontology Representation - Design Patterns and Ontologies that Make Sense*
2009-14 Maksym Korotkiy (VU), *From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)*
2009-13 Steven de Jong (UM), *Fairness in Multi-Agent Systems*
2009-12 Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin), *Operating Guidelines for Services*
2009-11 Alexander Boer (UVA), *Legal Theory, Sources of Law & the Semantic Web*
2009-10 Jan Wielemaker (UVA), *Logic programming for knowledge-intensive interactive applications*
2009-09 Benjamin Kanagwa (RUN), *Design, Discovery and Construction of Service-oriented Systems*
2009-08 Volker Nannen (VU), *Evolutionary Agent-Based Policy Analysis in Dynamic Environments*
2009-07 Ronald Poppe (UT), *Discriminative Vision-Based Recovery and Recognition of Human Motion*

2009-06 Muhammad Subianto (UU), *Understanding Classification*
2009-05 Sietse Overbeek (RUN), *Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality*
2009-04 Josephine Nabukenya (RUN), *Improving the Quality of Organisational Policy Making using Collaboration Engineering*
2009-03 Hans Stol (UvT), *A Framework for Evidence-based Policy Making Using IT*
2009-02 Willem Robert van Hage (VU), *Evaluating Ontology-Alignment Techniques*
2009-01 Rasa Jurgelenaite (RUN), *Symmetric Causal Independence Models*
2008-35 Ben Torben Nielsen (UvT), *Dendritic morphologies: function shapes structure*
2008-34 Jeroen de Knijf (UU), *Studies in Frequent Tree Mining*
2008-33 Frank Terpstra (UVA), *Scientific Workflow Design; theoretical and practical issues*
2008-32 Trung H. Bui (UT), *Toward Affective Dialogue Management using Partially Observable Markov Decision Processes*
2008-31 Loes Braun (UM), *Pro-Active Medical Information Retrieval*
2008-30 Wouter van Atteveldt (VU), *Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content*
2008-29 Dennis Reidsma (UT), *Annotations and Subjective Machines – Of Annotators, Embodied Agents, Users, and Other Humans*
2008-28 Ildiko Flesch (RUN), *On the Use of Independence Relations in Bayesian Networks*
2008-27 Hubert Vogten (OU), *Design and Implementation Strategies for IMS Learning Design*
2008-26 Marijn Huijbregts (UT), *Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled*
2008-25 Geert Jonker (UU), *Efficient and Equitable Exchange in Air Traffic Management Plan Repair using Spender-signed Currency*
2008-24 Zharko Aleksovski (VU), *Using background knowledge in ontology matching*
2008-23 Stefan Visscher (UU), *Bayesian network models for the management of ventilator-associated pneumonia*
2008-22 Henk Koning (UU), *Communication of IT-Architecture*
2008-21 Krisztian Balog (UVA), *People Search in the Enterprise*
2008-20 Rex Arendsen (UVA), *Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven*
2008-19 Henning Rode (UT), *From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search*
2008-18 Guido de Croon (UM), *Adaptive Active Vision*
2008-17 Martin Op 't Land (TUD), *Applying Architecture and Ontology to the Splitting and Allying of Enterprises*
2008-16 Henriëtte van Vugt (VU), *Embodied agents from a user's perspective*
2008-15 Martijn van Otterlo (UT), *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains*
2008-14 Arthur van Bunningen (UT), *Context-Aware Querying; Better Answers with Less Effort*
2008-13 Caterina Carraciolo (UVA), *Topic Driven Access to Scientific Handbooks*
2008-12 József Farkas (RUN), *A Semiotically Oriented Cognitive Model of Knowledge Representation*
2008-11 Vera Kartseva (VU), *Designing Controls for Network Organizations: A Value-Based Approach*
2008-10 Wauter Bosma (UT), *Discourse oriented summarization*
2008-09 Christof van Nimwegen (UU), *The paradox of the guided user: assistance can be counter-effective*
2008-08 Janneke Bolt (UU), *Bayesian Networks: Aspects of Approximate Inference*
2008-07 Peter van Rosmalen (OU), *Supporting the tutor in the design and support of adaptive e-learning*
2008-06 Arjen Hommersom (RUN), *On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective*
2008-05 Bela Mutschler (UT), *Modeling and simulating causal dependencies on process-aware information systems from a cost perspective*
2008-04 Ander de Keijzer (UT), *Management of Uncertain Data – towards unattended integration*
2008-03 Vera Hollink (UVA), *Optimizing hierarchical menus: a usage-based approach*
2008-02 Alexei Sharpanskykh (VU), *On Computer-Aided Methods for Modeling and Analysis of Organizations*
2008-01 Katalin Boer-Sorbán (EUR), *Agent-Based Simulation of Financial Markets: A modular, continuous-time approach*
2007-25 Joost Schalken (VU), *Empirical Investigations in Software Process Improvement*
2007-24 Georgina Ramírez Camps (CWI), *Structural Features in XML Retrieval*
2007-23 Peter Barna (TUE), *Specification of Application Logic in Web Information Systems*
2007-22 Zlatko Zlatev (UT), *Goal-oriented design of value and process models from patterns*
2007-21 Karianne Vermaas (UU), *Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005*
2007-20 Slinger Jansen (UU), *Customer Configuration Updating in a Software Supply Network*
2007-19 David Levy (UM), *Intimate relationships with artificial partners*
2007-18 Bart Orriëns (UvT), *On the development an management of adaptive business collaborations*
2007-17 Theodore Charitos (UU), *Reasoning with Dynamic Networks in Practice*
2007-16 Davide Grossi (UU), *Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems*
2007-15 Joyca Lacroix (UM), *NIM: a Situated Computational Memory Model*
2007-14 Niek Bergboer (UM), *Context-Based Image Analysis*
2007-13 Rutger Rienks (UT), *Meetings in Smart Environments; Implications of Progressing Technology*
2007-12 Marcel van Gerven (RUN), *Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty*
2007-11 Natalia Stash (TUE), *Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System*
2007-10 Huib Aldewereld (UU), *Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols*
2007-09 David Mobach (VU), *Agent-Based Mediated Service Negotiation*
2007-08 Mark Hoogendoorn (VU), *Modeling of Change in Multi-Agent Organizations*
2007-07 Nataša Jovanović (UT), *To Whom It May Concern – Addressee Identification in Face-to-Face Meetings*
2007-06 Gilad Mishne (UVA), *Applied Text Analytics for Blogs*

2007-05 Bart Schermer (UL), *Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance*
2007-04 Jurriaan van Diggelen (UU), *Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach*
2007-03 Peter Mika (VU), *Social Networks and the Semantic Web*
2007-02 Wouter Teepe (RUG), *Reconciling Information Exchange and Confidentiality: A Formal Approach*
2007-01 Kees Leune (UvT), *Access Control and Service-Oriented Architectures*
2006-28 Börkur Sigurbjörnsson (UVA), *Focused Information Access using XML Element Retrieval*
2006-27 Stefano Bocconi (CWI), *Vox Populi: generating video documentaries from semantically annotated media repositories*
2006-26 Vojkan Mihajlović (UT), *Score Region Algebra: A Flexible Framework for Structured Information Retrieval*
2006-25 Madalina Drugan (UU), *Conditional log-likelihood MDL and Evolutionary MCMC*
2006-24 Laura Hollink (VU), *Semantic Annotation for Retrieval of Visual Resources*
2006-23 Ion Juvina (UU), *Development of Cognitive Model for Navigating on the Web*
2006-22 Paul de Vrieze (RUN), *Fundaments of Adaptive Personalisation*
2006-21 Bas van Gils (RUN), *Aptness on the Web*
2006-20 Marina Velikova (UvT), *Monotone models for prediction in data mining*
2006-19 Birna van Riemsdijk (UU), *Cognitive Agent Programming: A Semantic Approach*
2006-18 Valentin Zhizhkun (UVA), *Graph transformation for Natural Language Processing*
2006-17 Stacey Nagata (UU), *User Assistance for Multitasking with Interruptions on a Mobile Device*
2006-16 Carsten Riggelsen (UU), *Approximation Methods for Efficient Learning of Bayesian Networks*
2006-15 Rainer Malik (UU), *CONAN: Text Mining in the Biomedical Domain*
2006-14 Johan Hoorn (VU), *Software Requirements: Update, Upgrade, Redesign – towards a Theory of Requirements Change*
2006-13 Henk-Jan Lebbink (UU), *Dialogue and Decision Games for Information Exchanging Agents*
2006-12 Bert Bongers (VU), *Interactivation – Towards an ecology of people, our technological environment, and the arts*
2006-11 Joeri van Ruth (UT), *Flattening Queries over Nested Data Types*
2006-10 Ronny Siebes (VU), *Semantic Routing in Peer-to-Peer Systems*
2006-09 Mohamed Wahdan (UM), *Automatic Formulation of the Auditor's Opinion*
2006-08 Eelco Herder (UT), *Forward, Back and Home Again – Analyzing User Behavior on the Web*
2006-07 Marko Smiljanic (UT), *XML schema matching – balancing efficiency and effectiveness by means of clustering*
2006-06 Ziv Baida (VU), *Software-aided Service Bundling – Intelligent Methods & Tools for Graphical Service Modeling*
2006-05 Cees Pierik (UU), *Validation Techniques for Object-Oriented Proof Outlines*
2006-04 Marta Sabou (VU), *Building Web Service Ontologies*
2006-03 Noor Christoph (UVA), *The role of metacognitive skills in learning to solve problems*
2006-02 Cristina Chisalita (VU), *Contextual issues in the design and use of information technology in organizations*
2006-01 Samuil Angelov (TUE), *Foundations of B2B Electronic Contracting*

2005-21 Wijnand Derks (UT), *Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics*
2005-20 Cristina Coteanu (UL), *Cyber Consumer Law, State of the Art and Perspectives*
2005-19 Michel van Dartel (UM), *Situated Representation*
2005-18 Danielle Sent (UU), *Test-selection strategies for probabilistic networks*
2005-17 Boris Shishkov (TUD), *Software Specification Based on Re-usable Business Components*
2005-16 Joris Graaumans (UU), *Usability of XML Query Languages*
2005-15 Tibor Bosse (VU), *Analysis of the Dynamics of Cognitive Processes*
2005-14 Borys Omelayenko (VU), *Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics*
2005-13 Fred Hamburg (UL), *Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen*
2005-12 Csaba Boer (EUR), *Distributed Simulation in Industry*
2005-11 Elth Ogston (VU), *Agent Based Matchmaking and Clustering – A Decentralized Approach to Search*
2005-10 Anders Bouwer (UVA), *Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments*
2005-09 Jeen Broekstra (VU), *Storage, Querying and Inferencing for Semantic Web Languages*
2005-08 Richard Vdovjak (TUE), *A Model-driven Approach for Building Distributed Ontology-based Web Applications*
2005-07 Flavius Frasincar (TUE), *Hypermedia Presentation Generation for Semantic Web Information Systems*
2005-06 Pieter Spronck (UM), *Adaptive Game AI*
2005-05 Gabriel Infante-Lopez (UVA), *Two-Level Probabilistic Grammars for Natural Language Parsing*
2005-04 Nirvana Meratnia (UT), *Towards Database Support for Moving Object data*
2005-03 Franc Grootjen (RUN), *A Pragmatic Approach to the Conceptualisation of Language*
2005-02 Erik van der Werf (UM)), *AI techniques for the game of Go*
2005-01 Floor Verdenius (UVA), *Methodological Aspects of Designing Induction-Based Applications*
2004-20 Madelon Evers (Nyenrode), *Learning from Design: facilitating multidisciplinary design teams*
2004-19 Thijs Westerveld (UT), *Using generative probabilistic models for multimedia retrieval*
2004-18 Vania Bessa Machado (UvA), *Supporting the Construction of Qualitative Knowledge Models*
2004-17 Mark Winands (UM), *Informed Search in Complex Games*
2004-16 Federico Divina (VU), *Hybrid Genetic Relational Search for Inductive Learning*
2004-15 Arno Knobbe (UU), *Multi-Relational Data Mining*
2004-14 Paul Harrenstein (UU), *Logic in Conflict. Logical Explorations in Strategic Equilibrium*
2004-13 Wojciech Jamroga (UT), *Using Multiple Models of Reality: On Agents who Know how to Play*
2004-12 The Duy Bui (UT), *Creating emotions and facial expressions for embodied agents*
2004-11 Michel Klein (VU), *Change Management for Distributed Ontologies*
2004-10 Suzanne Kabel (UVA), *Knowledge-rich indexing of learning-objects*
2004-09 Martin Caminada (VU), *For the Sake of the Argument; explorations into argument-based reasoning*

2004-08 Joop Verbeek (UM), *Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise*
2004-07 Elise Boltjes (UM), *Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes*
2004-06 Bart-Jan Hommes (TUD), *The Evaluation of Business Process Modeling Techniques*
2004-05 Viara Popova (EUR), *Knowledge discovery and monotonicity*
2004-04 Chris van Aart (UVA), *Organizational Principles for Multi-Agent Architectures*
2004-03 Perry Groot (VU), *A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving*
2004-02 Lai Xu (UvT), *Monitoring Multi-party Contracts for E-business*
2004-01 Virginia Dignum (UU), *A Model for Organizational Interaction: Based on Agents, Founded in Logic*
2003-18 Levente Kocsis (UM), *Learning Search Decisions*
2003-17 David Jansen (UT), *Extensions of Statecharts with Probability, Time, and Stochastic Timing*
2003-16 Menzo Windhouwer (CWI), *Feature Grammar Systems – Incremental Maintenance of Indexes to Digital Media Warehouses*
2003-15 Mathijs de Weerdt (TUD), *Plan Merging in Multi-Agent Systems*
2003-14 Stijn Hoppenbrouwers (KUN), *Freezing Language: Conceptualisation Processes across ICT-Supported Organisations*
2003-13 Jeroen Donkers (UM), *Nosce Hostem – Searching with Opponent Models*
2003-12 Roeland Ordelman (UT), *Dutch speech recognition in multimedia information retrieval*
2003-11 Simon Keizer (UT), *Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks*
2003-10 Andreas Lincke (UvT), *Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture*
2003-09 Rens Kortmann (UM), *The resolution of visually guided behaviour*
2003-08 Yongping Ran (UM), *Repair Based Scheduling*
2003-07 Machiel Jansen (UvA), *Formal Explorations of Knowledge Intensive Tasks*
2003-06 Boris van Schooten (UT), *Development and specification of virtual environments*
2003-05 Jos Lehmann (UVA), *Causation in Artificial Intelligence and Law – A modelling approach*
2003-04 Milan Petković (UT), *Content-Based Video Retrieval Supported by Database Technology*
2003-03 Martijn Schuemie (TUD), *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*
2003-02 Jan Broersen (VU), *Modal Action Logics for Reasoning About Reactive Systems*
2003-01 Heiner Stuckenschmidt (VU), *Ontology-Based Information Sharing in Weakly Structured Environments*
2002-17 Stefan Manegold (UVA), *Understanding, Modeling, and Improving Main-Memory Database Performance*
2002-16 Pieter van Langen (VU), *The Anatomy of Design: Foundations, Models and Applications*
2002-15 Rik Eshuis (UT), *Semantics and Verification of UML Activity Diagrams for Workflow Modelling*
2002-14 Wieke de Vries (UU), *Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems*
2002-13 Hongjing Wu (TUE), *A Reference Architecture for Adaptive Hypermedia Applications*
2002-12 Albrecht Schmidt (Uva), *Processing XML in Database Systems*
2002-11 Wouter C.A. Wijngaards (VU), *Agent Based Modelling of Dynamics: Biological and Organisational Applications*
2002-10 Brian Sheppard (UM), *Towards Perfect Play of Scrabble*
2002-09 Willem-Jan van den Heuvel (KUB), *Integrating Modern Business Applications with Objectified Legacy Systems*
2002-08 Jaap Gordijn (VU), *Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas*
2002-07 Peter Boncz (CWI), *Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications*
2002-06 Laurens Mommers (UL), *Applied legal epistemology; Building a knowledge-based ontology of the legal domain*
2002-05 Radu Serban (VU), *The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents*
2002-04 Juan Roberto Castelo Valdueza (UU), *The Discrete Acyclic Digraph Markov Model in Data Mining*
2002-03 Henk Ernst Blok (UT), *Database Optimization Aspects for Information Retrieval*
2002-02 Roelof van Zwol (UT), *Modelling and searching web-based document collections*
2002-01 Nico Lassing (VU), *Architecture-Level Modifiability Analysis*
2001-11 Tom M. van Engers (VUA), *Knowledge Management: The Role of Mental Models in Business Systems Design*
2001-10 Maarten Sierhuis (UvA), *Modeling and Simulating Work Practice BRAHMS: a multiagent modeling and simulation language for work practice analysis and design*
2001-09 Pieter Jan 't Hoen (RUL), *Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes*
2001-08 Pascal van Eck (VU), *A Compositional Semantic Structure for Multi-Agent Systems Dynamics*
2001-07 Bastiaan Schonhage (VU), *Diva: Architectural Perspectives on Information Visualization*
2001-06 Martijn van Welie (VU), *Task-based User Interface Design*
2001-05 Jacco van Ossenbruggen (VU), *Processing Structured Hypermedia: A Matter of Style*
2001-04 Evgueni Smirnov (UM), *Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets*
2001-03 Maarten van Someren (UvA), *Learning as problem solving*
2001-02 Koen Hindriks (UU), *Agent Programming Languages: Programming with Mental Models*
2001-01 Silja Renooij (UU), *Qualitative Approaches to Quantifying Probabilistic Networks*
2000-11 Jonas Karlsson (CWI), *Scalable Distributed Data Structures for Database Management*
2000-10 Niels Nes (CWI), *Image Database Management System Design Considerations, Algorithms and Architecture*
2000-09 Florian Waas (CWI), *Principles of Probabilistic Query Optimization*
2000-08 Veerle Coupé (EUR), *Sensitivity Analyis of Decision-Theoretic Networks*
2000-07 Niels Peek (UU), *Decision-theoretic Planning of Clinical Patient Management*
2000-06 Rogier van Eijk (UU), *Programming Languages for Agent Communication*
2000-05 Ruud van der Pol (UM), *Knowledge-based Query Formulation in Information Retrieval*
2000-04 Geert de Haan (VU), *ETAG, A Formal Model of Competence Knowledge for User Interface Design*
2000-03 Carolien M.T. Metselaar (UVA), *Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectief*
2000-02 Koen Holtman (TUE), *Prototyping of CMS Storage Management*

2000-01 Frank Niessink (VU), *Perspectives on Improving Software Maintenance*

1999-08 Jacques H.J. Lenting (UM), *Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation*

1999-07 David Spelt (UT), *Verification support for object database design*

1999-06 Niek J.E. Wijngaards (VU), *Re-design of compositional systems*

1999-05 Aldo de Moor (KUB), *Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems*

1999-04 Jacques Penders (UM), *The practical Art of Moving Physical Objects*

1999-03 Don Beal (UM), *The Nature of Minimax Search*

1999-02 Rob Potharst (EUR), *Classification using decision trees and neural nets*

1999-01 Mark Sloof (VU), *Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products*

1998-05 E.W. Oskamp (RUL), *Computerondersteuning bij Straftoemeting*

1998-04 Dennis Breuker (UM), *Memory versus Search in Games*

1998-03 Ans Steuten (TUD), *A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective*

1998-02 Floris Wiesman (UM), *Information Retrieval by Graphically Browsing Meta-Information*

1998-01 Johan van den Akker (CWI), *DEGAS – An Active, Temporal Database of Autonomous Objects*