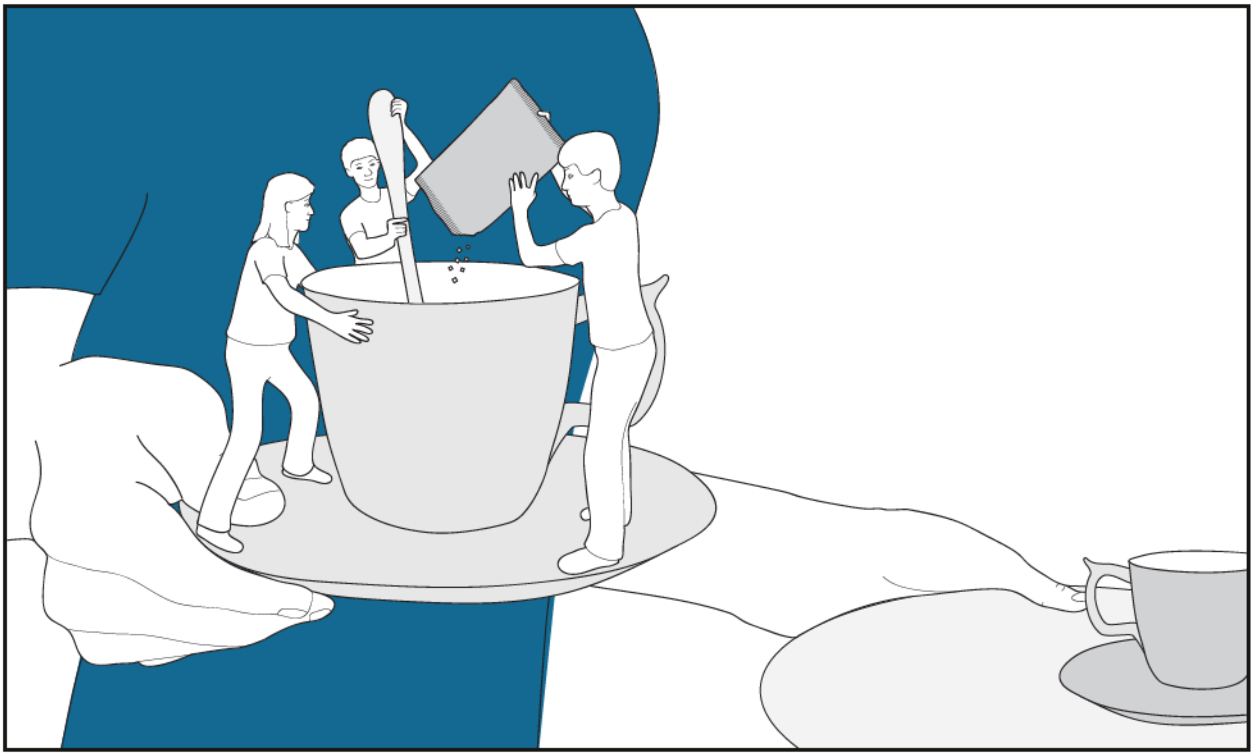
Alexander Kulik

User Interfaces for Cooperation



Bauhaus-Universität
Weimar

User Interfaces for Cooperation

Von der Fakultät Medien der Bauhaus-Universität Weimar zur Erlangung des akademischen Grades Doctor rerum naturalium (Dr. rer. nat.), genehmigte Dissertation von

Dipl. Des. Alexander Kulik
geboren am 28.12.1979 in Dresden

1. Gutachter: Prof. Dr. Bernd Fröhlich

2. Gutachter: Prof. Victoria Interrante, Ph.D.

3. Gutachter: Prof. Sue Cobb, Ph.D.

Tag der mündlichen Prüfung: 11.10.2016

Weimar 2016

To my parents

Contents

Table of Contents iv

Abstract xiii

Überblick xv

Acknowledgments xvii

Publications xxv

I Background and Motivation 1

1 Introduction 3

1.1 Specialization and Combination of Tools 4

1.1.1 Task-Specific Tools 5

1.1.2 Interface Adaptation 6

1.1.3 Virtual Interaction Widgets 7

1.1.4 Specialized Input Hardware 7

1.2 Towards Cooperative User Interfaces 8

1.2.1 Terminology 9

2 Performance Measures in Direct Manipulation Tasks 11

2.1 2D-Pointing Benchmarks 15

2.2 The Information Capacity of Full Body Motion 17

2.3 3D-Pointing Benchmarks 19

2.4 3D-Rotation Benchmarks 21

2.5 Object Manipulation Benchmarks 22

2.6 Conclusion 24

3 Integrality and Separability 27

3.1 Attribute Interrelations in Perceptual Processing 28

3.2 Integrality and Separability in Manipulation Tasks 32

3.2.1 Implications from Human Motor Control 35

3.2.2 Integral Attribute Manipulation with Separable Input 37

3.2.3 Integrality and Separability in Multitouch Interfaces 40

3.3 Conclusion 42

4 Cooperative Action and Movement Coordination 45

4.1 Synergies 46

4.2 Hierarchical Control 50

4.3 Bilateral Coordination 50

4.3.1 Bilateral Movement Synchronization 51

4.3.2 Synchronization through Perception-Action Coupling 53

4.4 Interpersonal Coordination 55

4.4.1 Joint Action . . . 57
4.5 Conclusion . . . 59

5 Related Work on Cooperative User Interfaces 61

5.1 Bimanual User Interfaces . . . 61
5.1.1 Touch Typing . . . 62
5.1.2 Bimanual Spatial Manipulation . . . 63
5.2 Multimodal Interaction . . . 65
5.2.1 Hand and Voice . . . 66
5.2.2 Gaze and Hand . . . 68
5.3 Multi-User Cooperation . . . 69
5.3.1 Manipulation Conflicts . . . 71
5.3.2 Cooperative Object Manipulation . . . 73
5.4 Conclusion . . . 76

6 Cooperation Patterns and Requirements 79

6.1 Cooperation Patterns . . . 79
6.1.1 Specialization . . . 80
6.1.2 Equivalence . . . 81
6.1.3 Confirmation and Intensification (Redundancy) . . . 82
6.1.4 Complementarity . . . 84
6.1.5 Transfer . . . 85
6.2 Workspace Requirements for Cooperation . . . 86
6.3 Levels of Autonomy . . . 88

6.4 Conclusion 90

II Approach and Experiments 93

7 The Design and Evaluation of Cooperative User Interfaces 95
7.1 Research Through Design 96
7.2 Research Program and Working Hypotheses 98
8 Bimanual Cooperation with Desktop 3D Input Devices 101
8.1 Introduction 102
8.2 Related Work 103
8.3 Desktop-Based 3D Interaction 104
8.4 Desktop-Based 3D Input Devices 105
8.5 Spatial Orientation Study 106
8.5.1 Task 107
8.5.2 Apparatus 109
8.5.3 Hypothesis 109
8.5.4 Design and Procedure 110
8.5.5 Participants 110
8.5.6 Results and Discussion 111
8.6 Motion Parallax Study 114
8.6.1 Task 114
8.6.2 Apparatus 116
8.6.3 Hypothesis 116

8.6.4 Design and Procedure 116
8.6.5 Participants 117
8.6.6 Results and Discussion 117
8.7 Conclusions and Future Work 119
Acknowledgements 119

9 Hold-and-Move – A Bimanual Cooperation Pattern 121
9.1 Introduction 122
9.2 Related Work 123
9.3 The Timing of Touch Gestures 124
9.3.1 Experimental Setup 124
9.3.2 Recorded Time Intervals 127
9.3.3 Design Considerations 128
9.4 The Hold-and-Move Input Pattern 129
9.5 User Study 131
9.5.1 Experimental Tasks 131
9.5.2 Participants 133
9.5.3 Design and Procedure 133
9.5.4 Hypotheses 133
9.5.5 Results 133
9.5.6 Error Analysis 135
9.5.7 Hold-Time Analysis 137
9.5.8 Discussion 138
9.6 Conclusions and Future Work 139

Acknowledgements 139

10 Multi-User Virtual Reality 141

10.1 Introduction 142

10.2 Related Work 143

10.3 Synchronized 12-View Projector Array 146

10.3.1 The 360Hz Projector 148

10.3.2 Video-Multiplexing and Geometric Alignment 149

10.4 Intelligent Shutter Glasses 150

10.5 Augmented Group Navigation 153

10.6 Results and Evaluation 158

10.6.1 Technical Evaluation 159

10.6.2 User Feedback 159

10.6.3 User Study 161

10.7 Conclusions and Future Work 163

Acknowledgements 165

III Further Examples and Conclusions 167

11 Case studies 169

11.1 Workspace Coherence in Remote Collaboration 170

11.1.1 The Effectiveness of Gestural Communication 171

11.1.2 Subjective Workspace Awareness 172

11.2 Territoriality in Immersive Virtual Environments 173

11.2.1 Visibility in Dense 3D Environments 174

11.2.2 Separate Views and Shared Perspectives 177

11.3 Complementary Tools for Cooperative 3D Interaction 182

11.3.1 Spatial References for Manipulation 183

11.3.2 Complementary Tools for Cooperative 3D Object Assembly . . 185

12 Conclusions and Future Work **189**

12.1 Thesis Contributions 192

12.2 Follow-Up Research 193

Bibliography **197**

Abstract

This thesis suggests cooperation as a design paradigm for human-computer interaction. The basic idea is that the synergistic co-operation of interfaces through concurrent user activities enables increased interaction fluency and expressiveness. This applies to bimanual interaction and multi-finger input, e.g., touch typing, as well as the collaboration of multiple users. Cooperative user interfaces offer more interaction flexibility and expressivity for single and multiple users.

Part I of this thesis analyzes the state of the art in user interface design. It explores limitations of common approaches and reveals the crucial role of cooperative action in several established user interfaces and research prototypes. A review of related research in psychology and human-computer interaction offers insights to the cognitive, behavioral, and ergonomic foundations of cooperative user interfaces. Moreover, this thesis suggests a broad applicability of generic cooperation patterns and contributes three high-level design principles.

Part II presents three experiments towards cooperative user interfaces in detail. A study on desktop-based 3D input devices, explores fundamental benefits of cooperative bimanual input and the impact of interface design on bimanual cooperative behavior. A novel interaction technique for multitouch devices is presented that follows the paradigm of cooperative user interfaces and demonstrates advantages over the status quo. Finally, this thesis introduces a fundamentally new display technology that provides up to six users with their individual perspectives of a shared 3D environment. The system creates new possibilities for the cooperative interaction of multiple users.

Part III of this thesis builds on the research results described in Part II, in particular, the multi-user 3D display system. A series of case studies in the field of collaborative virtual reality provides exemplary evidence for the relevance and applicability of the suggested design principles.

Überblick

Die vorliegende Arbeit betrachtet Kooperation als Gestaltungsparadigma für Mensch-Maschine Schnittstellen. Dabei geht es um Kooperation im Sinne paralleler Aktivitäten und deren synergetischer Kombination mit dem Ziel einer flüssigen und effektiven Computerarbeit. Dieses Interaktionsmuster ist für zweihändige Eingaben und die Nutzung mehrerer Finger, z.B. beim Maschinenschreiben, genauso anwendbar wie für die Zusammenarbeit mehrerer Nutzer. Kooperative Benutzungsschnittstellen bieten Einzelpersonen sowie Gruppen von Nutzern mehr Flexibilität und Ausdrucksmöglichkeiten.

Teil I dieser Arbeit betrachtet den Stand von Forschung und Technik zu diesem Thema. Dabei werden Limitierungen etablierter Benutzungsschnittstellen untersucht als auch das Potential und die Bedeutung kooperativer Interaktion. Auf Grundlage von Forschungsergebnissen aus der Psychologie, den Bewegungswissenschaften und der Forschung zu Mensch-Maschine Schnittstellen werden kognitive und ergonomische Grundlagen kooperativer Benutzungsschnittstellen abgeleitet. Darüber hinaus werden generische Kooperationsmuster diskutiert und die Anforderungen an kooperative Benutzungsschnittstellen in drei Gestaltungsprinzipien zusammengefasst.

Teil II dieser Arbeit stellt drei Forschungsarbeiten zur Entwicklung und Untersuchung kooperativer Benutzungsschnittstellen vor. In Kapitel 8 wird zweihändige Kooperation am Beispiel tischbasierter 3D Eingabegeräte untersucht. Kapitel 9 stellt eine neue Multitouch Interaktionstechnik vor, die dem Paradigma kooperativer Benutzungsschnittstellen folgt und klare Vorteile gegenüber einer etablierten Technik aufweist. Kapitel 10 präsentiert die Entwicklung und Untersuchung einer neuen 3D Projektionstechnologie, die bis zu sechs Personen individuelle Perspektiven auf eine gemeinsame virtuelle Umgebung bietet. Daraus ergeben sich völlig neue Möglichkeiten für die kooperative Interaktion mehrerer Nutzer mit dreidimensionalen Daten.

Teil III dieser Arbeit baut auf den Ergebnissen der in Teil II beschriebenen Experimente auf. Fallstudien aus dem Bereich der virtuellen Realität für mehrere Nutzer, belegen die Relevanz und Anwendbarkeit der vorgeschlagenen Gestaltungsprinzipien.

Acknowledgments

This is a thesis about user interfaces for cooperation. It is also a report on user interfaces that have been built in cooperation. I should not be writing these lines, if it weren't for the amazing company I enjoyed on the path towards this thesis. Perhaps, none of the developments and none of the research, described here, would have happened without the cooperation of those brilliant minds and skilled hands.

Successful cooperation builds on individual contributions of all participants – a commonplace that is also discussed in this thesis. The joint work with my colleagues is no exception. As a member of our research group, I contributed ideas, literature research, and experiments. I took responsibility for some of the many bits and pieces that constitute our experimental setups. In this thesis, I wrote up my own viewpoints and personal conclusions, but, every aspect of it reflects the results of fruitful cooperative work. The following paragraphs acknowledge the most influential contributions of others.

More than anyone else, it was Prof. Bernd Fröhlich, who paved the way for this thesis. He offered me to conduct this research under his supervision, although my degree in art and design did not even qualify for a Ph.D. in computer science. He believed in my curiosity and my ability to learn. He taught me some of the most essential skills. Not least, this includes programming and software engineering, which I studied in his lecture and practiced in exercises under the excellent supervision of Andreas Bernstein and André Schollmeyer. In this regard, I also thank the graduation committee who tailored an individual computer science curriculum for me as a precondition for submitting this thesis. I am grateful to all my teachers for the quality of their lectures and exercises. In particular, I thank Prof. Tom Gross for his lectures on HCI and CSCW and Prof. Anke Huckauf, who taught me the essentials of perceptual psychology and statistical evaluation methods.

Bernd Fröhlich did not only invite me to do exciting research, but he also created an inspiring and productive research environment. This includes many technical facilities, but, more importantly, an exceptional team. It is also evident that our work was always inspired by Bernd's long term vision of multi-user virtual reality. Back in the late nineties, he already worked on the necessary display technologies and interaction techniques (e.g. [4, 100]). Despite his agenda, he encouraged us to follow our own curiosity and find individual approaches to improving user interfaces. Thank you, Bernd, for sharing your ideas and your optimism! Thank you for all the inspiring discussions and your constructive critique! Your wit and wisdom is phenomenal. Whenever I felt overwhelmed by the obstacles along the way, you found the right words that encouraged me to appreciate the challenge.

The VR systems group at Bauhaus-Universität Weimar is a wonderful team to work with. I am grateful to my colleagues for the friendly and productive atmosphere and the open-minded exchange, we had and still have with one another. I had the pleasure to work with Stephan Beck, Andreas Bernstein, Daniel Fischer, Henning Gründl, Jan Hochstrate, André Kunert, Christopher Lux, Carl-Feofan Matthes, Hans-Friedrich Pabst, Patrick Riehmann, Andre Schollmeyer and Sebastian Thiele.

Among these colleagues, I want to emphasize those involved in our research on virtual reality and user interfaces. It was Jan Hochstrate, who introduced me to the topic of 3D user interfaces. He was also one of the inventors of the Globefish input device [98], which we used in the studies on bimanual cooperation with desktop-based 3D devices (Chapter 8). Daniel Fischer introduced the topic of multitouch tabletop interaction to our group. I appreciated the extensive discussions we had about general concepts of human-computer interaction and the topics of interaction fluency and immediacy in particular. Even after he left the group, we was heavily involved in the implementation of test applications for the Hold and Move interaction technique (Chapter 9). André Kunert and Stephan Beck were certainly my closest collaborators. We had exciting discussions about novel interactive technologies, the affordances of virtual phenomena, and the behavioral logics of an interactive mixed reality.

André Kunert facilitated our joint research through quasi-simultaneous translations of interaction concepts and software designs in operational Python code. Thank you, André, for sharing your knowledge and creativity! Your serenity and persistence make for an exceptional productivity. I would have lost my mind more than once over elusive bugs, if you had not resolved them calmly through consecutive attempts.

Stephan Beck created some of the most substantial software components of our experimental applications. In particular, he contributed the calibration routines for the multi-projector system described in Chapter 10 and he developed the complete pipeline for the real-time 3D acquisition, transfer, and rendering of 3D-video avatars presented in Section 11.1. Thank you, Stephan, for sharing your know-how and ingenuity! I admire your structured thinking and methodological rigor. Without your critical thinking and your profound understanding, we would have failed several critical challenges.

I also thank the many external collaborators with whom we have been working together in European research projects and exchange programs. This includes the par-

ticipants of the European research projects IMVIS, VR-Hyperspace, and 3D-Pitoti. The IMVIS project laid the foundations for the most exciting research related to this thesis. It allowed us to build the multi-user VR system described in Chapter 10. During VR-Hyperspace we extended this system with 3D telepresence technology (see Section 11.1) [26] and 3D-Pitoti gave us the opportunity to explore the design of 3D-user interfaces for the collaborative analysis of multi-scale 3D scanning data (see Section 11.2 and 11.3) [189].

Among our external collaborators I want to emphasize three participants of the IMVIS project: Roland Blach, Armin Zink and Stephen Palmer. Our joint project was critically important for this thesis. Roland Blach managed the IMVIS project. He also supervised my first experiments on multi-user 3D projection systems during an internship at Fraunhofer IAO in Stuttgart. During the project, he convinced us to use a radio link for the synchronization of the shutter glasses and implemented an initial demonstrator. Armin Zink with his company Digital Image[1] heavily modified commodity projectors to meet the requirements of our system. Moreover, I thank him for teaching me in the maintenance and repair of these devices. Stephen Palmer was working with LC-Tec[2] during that time and provided us with the custom liquid-crystal elements for our shutter glasses. Many thanks to Roland, Armin, and Stephen!

The projects VR-Hyperspace and 3D-Pitoti led to pioneering research and impressive technological developments in the involved research labs. The diversity of participants in both collaborations was a source of inspiration that eventually led to great results, but, it was also a challenge for the project coordination. I believe that the success of both projects must be largely attributed to the excellent management of Mirabelle D'Cruz, Sue Cobb, Harshada Patel, and Sally Shalloe from the University of Nottingham, UK. Our exchange was always productive and inspiring. If anyone, this team lives the spirit of cooperation. I am looking forward to future collaborations. Sadly without Sally, who passed away last year. We miss her.

Sue Cobb had a particular impact on this thesis. Towards the end of my Ph.D. research, when my motivation became declined during the final writeup, her academic interest renewed my enthusiasm. She is, without any doubt, a formidable expert in the topic. The concept of *complementarity*, which I emphasize as a fundamental building block of cooperative user interfaces, was pioneered several years ago in her research with colleagues from the University of Nottingham [27]. Sue was the first person to read this thesis in its entirety. She has proofread my writing for grammar and spelling errors and indicated where my argument was vague. Thank you, Sue, for your helpful comments!

During the course of my Ph.D. research, I had the opportunity to work with Martin Hachet at Inria Bordeaux, France, with François Guimbretière at Cornell University in Ithaca, USA and Ferran Argelaguet who came from Universitat Politècnica de

---

[1] http://www.digital-image.de

[2] http://www.lc-tec.se

Catalunya, Spain, to join our work in Weimar. I am grateful for the productive scientific exchange with these colleagues. The collaboration with Ferran was particularly relevant for my thesis. Thank you, Ferran, for exploring the issue of interpersonal occlusion with us (see Section 11.2)!

I had the privilege to supervise bachelor and master theses as well as training projects with very talented students. I enjoyed each of these collaborations, but I want to emphasize those that directly contributed to this thesis. Jan Dittrich was a student of media design, when he worked with us on multitouch interaction techniques. Our collaboration led to the development and evaluation of the Hold and Move interaction technique presented in Chapter 9. Florian Haupt built the first prototype of the Spheron group navigation device in a joint project with the faculty of industrial design. Sebastian Utzig developed the experimental 3D assembly application presented in Section 11.3 during his bachelor thesis. The bachelor thesis of Roman Reichel involved the implementation of the first functional control electronics for our multi-view 3D shutter glasses (Chapter 10). Sascha Gärtner and Susanne Günther supported the implementation of the experimental tasks for the evaluation of bimanual cooperation with desktop-based 3D input devices (Chapter 8). The bachelor theses of Nils Gründl and Onno Haak contributed to the iterative improvement of the Globfish device prototypes.

The research presented in this thesis was funded in part by the European Union under grant 217140 (project IMVIS), the EC FP7 projects 3D-PITOTI (ICT-2011-600545) and VR-HYPERSPACE (AAT-2011-RTD-1-285681). Our work was also supported by the German Federal Ministry of Education and Research (BMBF) under grant 03IP704 (project Intelligentes Lernen), grant 03IPT704X (project Big Data Analytics) and the Thuringian Ministry of Education and Cultural Affairs (TKM) under grant B514-08028 (project Visual Analytics in Engineering).

Above all, I thank my family for their love and support. Dziękuję rodzice, dziękuję Sylvia! Danke Maria und Eleonora! Danke Anette und Jörg! Thank you for your patience and your trust. You kept me going, when I was exhausted.

Erklärung über die Eigenständigkeit der Dissertation

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Teile der Arbeit die bereits Gegenstand von Prüfungsarbeiten waren, sind ebenfalls unmissverständlich gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich oder unentgeltlich geholfen:

Prof. Bernd Fröhlich:	Wissenschaftliche Betreuung der Arbeit
André Kunert:	Softwareentwicklungen für Experimente zu zweihändiger 3D Interaktion mit tischbasierten Eingabegeräten (Kapitel 8) sowie zu Mehrbenutzerinterkation in virtuellen Umgebungen (Kapitel 10, 11.2.2 und 11.3.1)
Stephan Beck:	Implementierung der Kalibrierungsroutinen für das in Kapitel 10 beschriebene Multi-Projektor System sowie Technologieentwicklung für die dreidimensionale Erfassung, und Rekonstruktion von Personen in Echtzeit für die Fallstudie zu “Workspace Coherence” in Telepräsenzsystemen (Kapitel 11.1).
Ferran Argelaguet:	Gemeinsame Konzeption und Auswertung der Experimente zur Sichtbarkeit von Interaktionsobjekten für mehrere Personen mit unterschiedlichen Perspektiven (Kapitel 11.2.1).

Jan Hochstrate, Susanne Günther und Sascha Gärtner:	Softwareentwicklung für Experimente zu zweihändiger 3D Interaktion mit tischbasierten Eingabegeräten (Kapitel 8)
Jan Dittrich:	Gemeinsame Konzeption der Hold-and-Move Interaktionstechnik für Multitouchgeräte (Kapitel 9.4) sowie Softwareentwicklungen für die Untersuchung von Aktivitätsmustern bei der Bedienung von Multitouchgeräten (Kapitel 9.3)
Felicitas Höbelt:	Softwareentwicklung für die Untersuchung von Aktivitätsmustern bei der Bedienung von Multitouchgeräten (Kapitel 9.3).
Daniel Fischer:	Auftragsentwicklung von Software für die Evaluierung der Hold-and-Move Interaktionstechnik (Kapitel 9.5)
Sebastian Utzig:	Implementierung des Versuchsaufbaus für die Fallstudie zu kollaborativer 3D Objektkonstruktion (Kapitel 11.3.2)
Sue Cobb:	Englischkorrektur

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlung- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Weimar, November 2016
Alexander Kulik

Declaration of Authorship

I hereby declare that this dissertation is entirely my own work except where otherwise indicated. I used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

The following people supported my PhD research and the writing of this thesis with specific contributions as described:

Prof. Bernd Fröhlich:	Scientific supervision
André Kunert:	Software developments for the experiments on bimanual 3D Interaction with desktop-based input devices (Chapter 8) and on multi-user interaction in virtual environments (Chapter 10, 11.2.2, and 11.3.1)
Stephan Beck:	Calibration routines for the multi-projector system described in Chapter 10 and the development of novel technologies for real-time 3D capturing and reconstruction of people for the case study on Workspace Coherence in telepresence applications (Chapter 11.1).
Ferran Argelaguet:	Joint design and evaluation of experiments on object visibility for multiple people with different perspectives (Chapter 11.2.1).
Jan Hochstrate, Susanne Günther, and Sascha Gärtner:	Software development for experiments on bimanual 3D Interaction with desktop-based input devices (Chapter 8)
Jan Dittrich:	Joint development of the Hold-and-Move interaction technique for multitouch devices (Chapter 9.4) and software developments for the analysis of temporal patterns in the operation of multitouch devices (Chapter 9.3)

Felicitas Höbelt:	Software development for the analysis of temporal patterns in the operation of multitouch devices (Chapter 9.3).
Daniel Fischer:	Implementation of software for the evaluation of the Hold-and-Move interaction technique (Chapter 9.5)
Sebastian Utzig:	Implementation of the experimental setup for the case study on collaborative 3D assembly design (Chapter 11.3.2)
Sue Cobb:	English proofreading

This thesis has never been submitted in the same or substantially similar version to any other examination office.

Weimar, November 2016
Alexander Kulik

List of Publications

During the course of this thesis I had the opportunity to work with many exceptional colleagues. Together we explored a variety of topics in the fields of human-computer interaction and 3D computer graphics and several results of our research were published. Three of these joint publications are core parts of this thesis. They are reprinted entirely in Part II. The described concepts and technical developments are joint work of the authors. Additional contributors are acknowledged at the end of the corresponding chapters. The author of this thesis was entirely responsible for the design and evaluation of all user studies reported in Part II. Several aspects of further co-authored publications are summarized in Part III. Figures taken from these publications have been indicated accordingly.

Publications included in Part II of this thesis

Kulik, A., Hochstrate, J., Kunert, A., Fröhlich, B.
The Influence of Input Device Characteristics on Spatial Perception in Desktop-Based 3D Applications
In Proceedings of the 2009 IEEE Symposium on 3D User Interfaces 2009 (3DUI 2009), pp. 59-66, 2009. DOI=http://dx.doi.org/10.1109/3DUI.2009.4811206
© 2009 IEEE, reprinted with permission in **Chapter 8**

Kulik, A., Dittrich, J., Fröhlich, B.
The Hold-and-Move Gesture for Multi-touch Interfaces
In Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12). pp. 49-58, September 2012.
DOI=http://dx.doi.org/10.1145/2371574.2371583
© 2012 ACM, reprinted with permission in **Chapter 9**

Kulik A., Kunert A., Beck S., Reichel R., Blach R., Zink A., Fröhlich B.
C1x6: A Stereoscopic Six-User Display for Co-located Collaboration in Shared Virtual Environments
In Proceedings of the 2011 SIGGRAPH Asia Conference (SA '11). ACM, New York, NY, USA, Article 188, 12 pages. DOI=http://dx.doi.org/10.1145/2024156.2024222
© 2011 ACM, reprinted with permission in **Chapter 10**.

Publications referenced in Part III in this thesis

Kunert, A., Kulik, A., Beck, S., Fröhlich B.
Photoportals: Shared References in Space and Time
In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14). ACM, New York, NY, USA, pp. 1388-1399, February 2014 DOI=http://dx.doi.org/10.1145/2531602.2531727
Referenced in **Section 11.2.2 and Section 11.3.1**.

Beck, S., Kunert, A., Kulik, A., Fröhlich B.
Immersive Group-to-Group Telepresence
IEEE Transactions on Visualization and Computer Graphics, 19(4) IEEE Computer Society, Los Alamitos, CA, USA, pp. 616-25, March 2013.
DOI=http://dx.doi.org/10.1109/TVCG.2013.33
Referenced in **Section 11.1**.

Argelaguet, F., Kunert, A., Kulik, A., Fröhlich B.
Improving Collaboration in Co-located Interactions with Show-Through Techniques
In Proceedings of IEEE Symposium on 3D User Interfaces IEEE Computer Society, Los Alamitos, CA, USA, pp. 55-62, March 2010.
DOI=http://dx.doi.org/10.1109/3DUI.2010.5444719
Referenced in **Section 11.2**.

Further related publications

Schlattmann, M., Yu, Y., Gründl, N., Bogen, M., Kulik, A., d'Angelo, D., Fröhlich, B., Klein, R.
User Awareness for Collaborative Multi-Touch Interaction
In Proceedings of the 8th International Conference on Computer Graphics Theory and Applications (GRAPP) Barcelona, Spain, pp. 359-366, February 2013

Ewerling, P., Kulik, A. Fröhlich, B.
Finger and hand detection for multi-touch interfaces based on maximally stable extremal regions
In Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces (ITS '12) ACM, New York, NY, USA, pp. 173-182, November 2012.
DOI=http://dx.doi.org/10.1145/2396636.2396663

Kulik, A., Kunert, A., Huckauf, A., Fröhlich, B.
The Groovepad: Ergonomic Integration of Isotonic and Elastic Input for Efficient Control of Complementary Subtasks
In Proceedings of APCHI 2012, The 10th Asia-Pacific Conference on Computer-Human Interaction. ACM, New York, NY, USA, pp. 28-31, August 2012.
DOI=http://dx.doi.org/10.1145/2350046.2350065

Argelaguet, F., Kulik, A., Kunert, A., Andujar, C., Fröhlich, B.
See-through Techniques for Referential Awareness in Collaborative Virtual Reality
International Journal of Human-Computer Studies (IJHCS), Volume 69, Issue 6, IEEE Computer Society, Los Alamitos, CA, USA, pp. 387-400, June 2011.
DOI=http://dx.doi.org/10.1016/j.ijhcs.2011.01.003

Kunert, A., Kulik, A., Lux, C., and Fröhlich, B.
Facilitating system control in ray-based interaction tasks
In Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology. ACM, New York, NY, pp. 183-186, November 2009.
DOI=http://dx.doi.org/10.1145/1643928.1643969

Kulik, A.
Building on Realism and Magic for Designing 3D Interaction Techniques
IEEE Computer Graphics and Applications, vol. 29, no. 6, IEEE Computer Society, Los Alamitos, CA, USA, pp. 22-33, Nov./Dec. 2009.
DOI=http://dx.doi.org/10.1109/MCG.2009.115

Kulik, A., Kunert, A., Lux, C., Fröhlich, B.
The Pie Slider: Combining Advantages of the Real and the Virtual Space
In Proceedings of the 10th international Symposium on Smart Graphics, LNCS vol. 5531, Springer-Verlag, Berlin, Heidelberg, pp. 93-104, 2009.
DOI=http://dx.doi.org/10.1007/978-3-642-02115-2_8

Kunert, A., Kulik, A., Huckauf, A., Fröhlich, B.
A comparison of tracking- and controller-based input for complex bimanual interaction in virtual environments
In Proceedings of the 13th Eurographics conference on Virtual Environments (EGVE'07), Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, pp. 43-52.
DOI=http://dx.doi.org/10.2312/EGVE/IPT_EGVE2007/043-052

Part I

Background and Motivation

Chapter 1

Introduction

> *"A self that is only differentiated – not integrated – may attain great individual accomplishments, but risks being mired in self-centered egotism. By the same token, a person who self is based exclusively on integration will be well connected and secure, but lack autonomous individuality. Only when a person invests equal amounts of psychic energy in these two processes and avoids both selfishness and conformity is the self likely to reflect complexity."*
>
> *—Mihaly Csikszentmihalyi [66]*

The driving applications of computer science have long been in the areas of data analysis and productivity, but applications for communication, information exchange, and entertainment have become at least equally relevant. Compared to the increased diversity of applications, computer interfaces have remained almost unchanged. They were generally designed for single users and only a small fraction of human skills were taken into account for the interaction with digital content and other users. The accessibility of powerful digital tools through a few button clicks of a single user is still a major motivation of interface design. Recently, the experience of users has gained more attention, but task efficiency remains an equally valid criterion for the quality of interfaces.

Efficient work with well suited tools can be very satisfying, while complex tools that fail to smoothly support simple subtasks are rather frustrating. Consider computer-aided design (CAD) as an example. The possibility to perfectly define geometrical relations with a few mouse clicks is overwhelming, but finding the most suitable operations in multi-level menu hierarchies is very time consuming and the 3D view adaptations that are necessary to see the relevant geometrical features are not well supported with the predominant 2D interaction paradigm of desktop computing. The experience of designing 3D shapes often involves more mode selections than geometrical manipulations. Kinesthetically, the process can hardly be distinguished from bookkeeping with spreadsheets. The interaction with current computer interfaces partially dissociates our minds from our own bodies and our immediate social environment. Djajadiningrat et al. argue that the neglected importance of bodily experiences and movement is not a very recent phenomenon, but one that roots in a longer history of industrial design [75].

We can observe a current trend to change this situation (e.g. [75, 158]). Novel sensor technologies and ever increasing computing power promise the development of more holistic user interfaces. Towards this objective, this thesis follows the idea that a diversity of simultaneously available interaction possibilities can increase the expressiveness of user input through a deeper involvement of cognitive and physical skills and at the same time encourage the cooperation of multiple users. Our everyday life is full of examples where cooperation improves both the efficiency and the experience of our activities. Peeling potatoes works better, when both hands are used together and the whole cooking process will be smoother, faster, and more enjoyable, if others contribute simultaneously to its success. Granted, that we do not use computers for peeling or cooking potatoes, but the smooth integration of multiple simultaneous user actions can be equally beneficial for computer applications that involve a continuous process of iterative changes and their evaluation, e.g., in visual analysis or the exploration of a design space.

1.1 Specialization and Combination of Tools

Many computing applications involve several fundamentally different operations like 2D pointing, 3D rotation, text entry, or (on the more complex side) volumetric selections. Just as mouse and keyboard nicely support combinations of 2D pointing and text entry tasks, further dedicated tools could be provided. The diversity of interface prototypes in research supports this idea. The combination of complementary tools can enable more complex actions with less individual effort from more involved actors.

Craftspeople have always equipped their workshops with a huge number of different tools. The enormous quantity of items we find in a common workshop is only partly related to different operations, but every type of tool must be available at various scales (Figure 1.1). Digital tools offer dynamic scaling and related adaptation techniques to adjust these properties on demand, hence a smaller number of devices appears to be necessary. In computer applications, one physical input device (e.g. the mouse) commonly serves for various tasks with exchangeable digital tool tips like pencils, brushes, or selection pointers. On the digital side of this amalgamation, application designers keep adding novel modalities to ever growing tool bars. The comparison with traditional workplaces, however, also indicates that the quantity of modes and digital widgets can be reduced, if the required functionalities can be achieved through the combination of multiple simultaneous input facilities. For example, drawing a straight line does not necessarily require another type of tool than drawing freeform shapes, but, the same pencil can be used with or without a ruler.

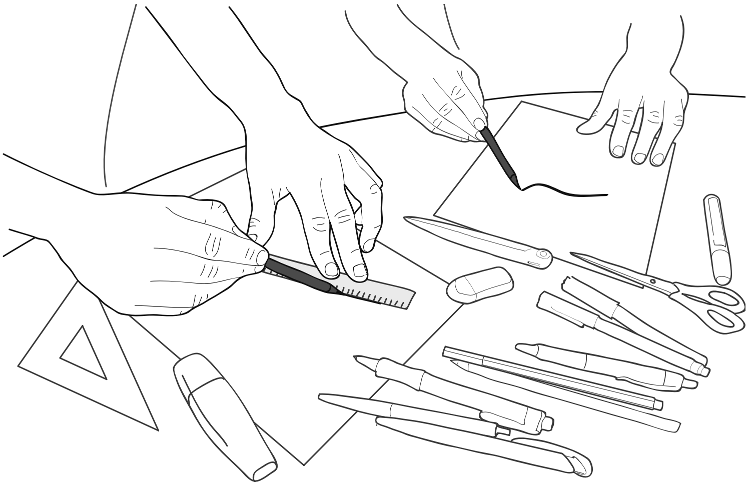


Figure 1.1: *Real workplaces are often crowded with tools. They can be used concurrently and they afford meaningful combinations for extended functionality.*

The effects of tool combinations are limited by physical constraints in traditional workshops. Computer applications, on the other hand, can potentially offer meaningful interpretations for any possible combination of inputs. Hinckley et al., for example, recently demonstrated an encouraging collection of functionalities that can be realized through such combined input of touch and pen [143]. The specialization of tools and their use in combination are fundamental principles of interaction design. Examples can be found everywhere, not least, in the realm of current computer interfaces. The following sections introduce some of the latter.

1.1.1 Task-Specific Tools

Based on an analysis of the benefits and drawbacks of various interface technologies and interaction techniques, we can choose the most suitable ones for the different requirements of text entry, object rotations, color adjustments, and many other tasks. The more we narrow down the specifications of the task, the more we can optimize for best performance. Practical applications, however, involve several such tasks with varying requirements. There are two approaches to deal with this diversity of demands. A generic interaction tool can be provided that serves the majority of tasks reasonably well or highly specialized instruments can be devised for each of them. The optimal balance between universality and differentiation always depends on the context of application.

Besides task correspondence, the general user performance in the manipulation of various interaction tools must also be considered. Direct manipulation seems to be most efficient if it is bound to physical support surfaces for motion input (see Chapter 2). Direct 3D manipulation of virtual objects lacks such physical support. Therefore, it is not surprising that 2D user interfaces are by far more established than their 3D counterparts – even for the specification of 3D transformations. Whether the performance benefits of 2D input outweigh the drawback of more indirect input mappings is open to debate and further research. Proponents of 3D input devices often argue with the integral operation of related attributes, benefits of proprioceptive feedback and the direct kinesthetic correspondence to the interaction task (e.g. [138, 231, 281, 329, 357]), while critique is often centered on lower input accuracy and higher fatigue (e.g. [31, 335]).

Be that as it may, people tend to move simultaneously through multiple dimensions and attribute spaces during coarse target approximation [21,131,190,242,356,357,381]. During fine grained parameter adjustments, instead, different degrees of freedom are often operated subsequently [21, 250, 251, 347]. For 3D object manipulation, the predominant 2D input paradigm can be considered a middle ground between both demands. However, it neither supports unconstrained direct manipulation for coarse approximation of 3D targets, nor does it offer implicit constraints to a single dimension for accurate placement. On-screen widgets are therefore used to further reduce the input to only a single axis.

1.1.2 Interface Adaptation

User interfaces can also be adapted dynamically to changing requirements, either implicitly or explicitly. The most successful examples of the first approach are adaptations of the transfer function based on motion velocity (e.g. [97,242]) and automatic object snapping to potential targets or related constraints (e.g. [25,36,37]). Intuitively this approach seems to offer the most potential if the adaptation builds on raw user input with many simultaneously operated degrees of freedom and applies reduction only when necessary towards the end of a placement task.

Alternatively, users can explicitly adapt parameters of the transfer function or chose among tools with different characteristics. The accessible presentation of multiple tools and settings, however, occupies valuable interaction space – either physical or virtual. Moreover, the required choice among multiple options can be detrimental as described by Hick's Law [132] and switching between them takes time. These issues can be more or less pronounced, depending on the type of involved tools and their arrangement. If not designed properly, the drawbacks can impair the benefits of dedicated task suitability. Ideally, the choice of tools, modes, and settings could be specified implicitly while focusing on an uninterrupted manipulation process. In future, brain-computer interfaces could be used to realize such implicit mode changes [320]

1.1.3 Virtual Interaction Widgets

Most computer interfaces offer only few physical input options, and provide additional functionality with on-screen widgets. These virtual tools can be operated with the same physical input. They can be displayed and arranged dynamically in relation to the operable parameters of selected objects, which facilitates the users' choice among them. Widgets can be used, for example, to translate 2D motion input from a pointer in screen space to other attribute spaces of an application, e.g., color or 3D position. The mapping often involves a reduction of the two-dimensional input to a single parameter, e.g. via a slider widget. Also the aforementioned mapping of 2D input to 3D object manipulations is often realized with on-screen widgets that support subsequent transformations on individual axes (see [302] for an overview).

Widgets are graphic representations of their functionalities, which makes them very versatile and comprehensible. Unfortunately, the available degrees of freedom can only be reduced and not increased. Tasks that involve multiple degrees of freedom must be operated in multiple steps. Consider the assembly of a complex 3D object from multiple parts using a mouse with digital manipulation widgets. For each object manipulation, users must operate translations and rotations along three spatial dimensions subsequently – each time involving the acquisition of the corresponding handle. It seems unlikely that this is the most effective interaction method.

1.1.4 Specialized Input Hardware

The provision of multiple specialized physical input devices can be beneficial for several reasons. Most importantly, the shape and weight of physical devices affords different types of action. For example, 2D input can be realized with direct touch, a pen, or a mouse device. Drawings require such 2D input, but the results differ strongly between the three technologies [373]. Furthermore, the design of physical input devices can accurately fit the type and the number of parameters that are to be operated integrally. Last but not least, physical devices support adjustments in various attribute spaces based on tactile and kinesthetic feedback. As a consequence, physical tools may require less visual monitoring for their operation than virtual interaction widgets [188, 362]. Users can even operate multiple physical input devices without losing their focus on higher-level aspects of the interaction task [137, 342]

As mentioned above, the simultaneous availability of multiple input devices affords their combination in various meaningful ways to achieve additional functionalities. Depending on the task at hand, their individual capabilities can be constrained or extended. The most established combination of computer interfaces is probably that of mouse and keyboard. The 2D pointing device enables direct manipulation while the keyboard provides symbolic input. Good interaction design exploits both interfaces for more efficient operation. We know several common patterns of their complementary use. When placing objects on a virtual canvas, for example, coarse approximation is rapidly achieved with a pointing device, but the direction keys of the keyboard facilitate fine-grained adjustments on individual axes.

1.2 Towards Cooperative User Interfaces

Apart from the above mentioned basic examples of input specialization and combination, HCI researchers have explored a variety of input devices and interaction styles that build on the synergistic cooperation of multiple parallel input streams. Chapter 5 reviews research in this realm with a particular focus on bimanual interaction, multimodal interaction, and interfaces that support the cooperation of multiple users. The actual usability of those novel interface developments cannot always be assessed from their isolated presentation in research papers. Comparisons with the multitude of alternatives are not always feasible. Chapter 2, therefore, reviews performance studies for direct manipulation interfaces and attempts to derive benchmark approximations for atomic operations.

The design of specialized interaction tools and their combinations must take into account which combinations or groupings of application parameters are useful and which should remain separated. From the perspective of cooperative action, this question extends to which degrees of freedom should be operated by a single actor and which can be beneficially distributed. Jacob et al. suggested to build such decisions on psychological research on the integrality and separability of perceptual attributes [157]. Chapter 3 reviews literature in this field and discusses the applicability of this perceptual typology to the design of interfaces for the simultaneous manipulation of multiple degrees of freedom.

Synergistic cooperation requires tight coordination. Research on human motor control has explored the coordination of multiple limbs of a single person. Sophisticated human motor skills like locomotion involve complex coordination patterns that result from extensive training during childhood. More recently, research on bilateral coordination and the joint action of multiple people, demonstrated several similarities between both cases [238, 283, 301, 303]. Therefore, user interfaces that are operated by multiple limbs, e.g. bimanually, and those for multi-user collaboration may have similar requirements. Chapter 4 discusses psychological research on synergistic cooperative action in human motor coordination, and specifically, in bilateral coordination and social interaction. Chapter 5 reviews related work in the design and evaluation of bimanual, multimodal, and collaborative user interfaces. Chapter 6 concludes Part I with a discussion of cooperation patterns and the proposal of design principles for cooperative user interfaces.

The second part of this thesis describes our design-oriented research program towards cooperative user interfaces and discusses the methodological pitfalls of this approach. Working hypotheses are defined that guided our experiments. Three selected experiments are presented in detail that demonstrate the successful applicability of cooperation as a design paradigm in very different application cases. Chapter 8 presents a formal user study on the performance benefits of bimanual cooperation in 3D orientation tasks and the effects of different input device characteristics. Chapter 9 presents a novel multitouch gesture that exploits temporal patterns of bimanual cooperation for implicit mode switching and proves usability benefits of the improved interaction fluency. Chapter 10 presents the design and evaluation of a novel display technology for the collocated collaboration of up to six users in an immersive virtual reality.

This thesis contributes novel interface developments and the evaluation of their usability. Moreover, it contributes design principles for cooperative user interfaces that have been derived from research results in the fields of cognitive psychology, motor control and HCI. The results can inform the design of future computer workplaces that encourage the cooperation of multiple users and a stronger involvement of human motor skills.

1.2.1 Terminology

Cooperative user interfaces are meant to support single users as well as multiple ones. The term *cooperation* does not only refer to interpersonal cooperation. It has also been established to describe cooperative actions of multiple human limbs (e.g. [116, 130, 138, 364]) and multimodal interaction (e.g. [61, 380]). *Cooperation*, as it is used in this thesis, refers to the synergistic, coordinated action of multiple entities that may be of various kinds. In statements about cooperation and coordination that apply to various different situations, e.g., bilateral coordination and multi-user collaboration, we use the term *actors* to represent these different types of cooperating entities.

In publications on the collaboration of multiple people, groups, and organizations, the terms cooperation, coordination, and collaboration are often used to specify three different levels of mutual interrelations. In Mattessich and Monsey's terminology, for example [225], cooperation refers to loose, informal relationships without mutual commitments. Everybody pursues only individual goals. Cooperation, is meant to occur spontaneously without the need for joint planning. Coordination, in this terminology, describes a much closer intertwining, where each actor still pursues only individual goals, but, concrete actions are adjusted to each other in order to minimize interference and increase the opportunities for synergistic cooperation. Collaboration is considered to be the strongest form of acting together. People collaborate in order to achieve goals they could not realize alone. This requires joint planning and the subordination of individual interests.

Our definition of these terms is slightly different. *Cooperation* refers to the concurrent operation of an interface by multiple entities/actors. This does not imply whether or not a common goal is pursued or if a goal can be specified at all. The term *coordination* describes the synchronization and alignment of actors, which we consider to be a prerequisite for successful cooperation. The term *collaboration* is only used in this thesis to specify interpersonal cooperation.

Chapter 2

Performance Measures in Direct Manipulation Tasks

With the success of graphical user interfaces, direct manipulation [316] became a prime paradigm in human computer interaction. Users can change the state of an application by manipulating objects on the screen. The comprehensibility of this interface concept has made computer applications usable for a wide range of users, many of which would struggle with written commands and responses only. However, several operations cannot be efficiently expressed through gestures and geometrical transformations. Information retrieval, for example, benefits from descriptive search terms and thus favors command-line interfaces [248]. However, almost all end-user applications apply direct manipulation for basic operations like view navigation and object selection. Moreover, direct manipulation of visually represented application content is a core feature of productivity software in computer-aided design, image editing, or information visualization.

Manipulation processes can be broken down into a sequence of movements towards intermediate or final target locations and/or orientations. Thus, the information capacity of user input, i.e., the potential rate of information throughput from the user to the system, is directly related to the accuracy and the number of such target acquisitions within a certain time. Continuous movements along a path, as required in tracing and steering tasks, can be modelled as a sequence of multiple intermediate target acquisitions [2, 207]. More complex tasks can be modelled as a sequence of actions that include physical, cognitive, and perceptual aspects which may partially overlap[1]. Besides cognitive processes, aimed movements constitute the most important subtasks of operating direct manipulation interfaces and, consequently, they are extensively studied in the HCI community.

Individual target acquisition tasks can be modelled with Fitts's law[2] [87]: a linear regression model with the intercept a and the slope coefficient b that relates the required movement time (MT) to the task's index of difficulty (ID). The latter is defined as the binary logarithm of the target's size-distance ratio (see equation 2.1). The logarithmic relation reflects that movements towards a known target involve a ballistic phase for coarse approximation. Further away targets are approached with a higher velocity, and thus the movement time does not increase linearly with the target distance. The smaller the target, the earlier this open-loop process must give way for closed-loop control with higher motion accuracy[3]. Fitts's original formulation of the model reads as follows:

$$MT = a + b * log_2(2D/W) \tag{2.1}$$

The model facilitates the comparison of various interfaces, if additional factors of user performance are otherwise controlled. The skills of test users or potential distractions and cognitive load imposed by of the test setting, for example, are not explicitly expressed in Fitts's law parameters. If these factors are not controlled, they mainly affect the intercept a of the regression model [377]. Moreover, the performance in aimed movement tasks strongly depends on the users' strategy. More than 30% faster task completion times have been observed if rapidity was favored over accuracy [214]. The side effect of such rapidity are larger deviations from the actual motion endpoint. MacKenzie suggested to compute the effective target width and distance from this actual distribution of hits which results in a comparable throughput measure that is independent from user strategies [214].

The use and formulation of Fitts's law is disputed in the research community. Several alternative formulas have been suggested for the computation of the index of difficulty (ID). The Shannon formulation that was suggested by MacKenzie [212] is currently the most popular in the HCI community. However, Drewes recently argued that Fitts's Law in its original form is a mathematical model of the target acquisition effort which should not be spoiled for the sake of a better fit with empirical data that

---

[1] Human interaction with computer interfaces can be analyzed according to the user's Goals, the available Operations, the Methods of their combination to achieve intermediate steps and the user's rules for the Selection of possible methods (GOMS [52]). This model supports the structural analysis of interaction sequences. The Keystroke Level Model is such a GOMS-based model that assumes a fixed time effort for elementary user actions with desktop workstations, i.e. keystrokes, aimed movements, and mental preparations [51], which allows the theoretical comparison of different interfaces based on the sum of required efforts.

[2] see [212] and [377] for extensive introductions to Fitts's law in HCI

[3] see [246] for a detailed analysis of movement phases

includes various sources of error [78]. The indices of difficulty discussed in the following literature review therefore correspond to Fitts's original formula (see above): the binary logarithm of target distance divided by half the target width[4]. IDs provided according to MacKenzie's Shannon formulation were recalculated based on available task descriptions. The differences rarely exceeded one bit.

Also for computing the index of performance (IP) or throughput[5] (in bits/s), several approaches have been suggested. Throughput is generally defined as the ratio of ID and MT [87, 321]. It is thus related to both, the intercept a and the slope coefficient b of the regression model. Shumin Zhai discussed the different effects of both factors on throughput computation [377]. He emphasized that the intercept a has a stronger effect for small IDs than for large ones. More reliable comparisons across various IDs may thus be possible if only the reciprocal of the slope coefficient b is considered. This approach, however, obscures constant overheads of a system, e.g., systematic activation latency of the input sensor.

The intercept a, also captures the user's reaction time, which strongly depends on the type of task. Fitts's original tapping task required repetitive movements between two fixed targets, hence a cyclic movement that required no or only little movement planning (Figure 2.1). Moreover, cyclic movements can be performed with less effort, if the target size allows for motion continuity. In a repetitive target acquisition task with low difficulty (linear slider with 10 cm target distance and about 3 mm target width) Guiard found that acceleration profiles resemble simple harmonic motion as modeled with Hooke's law for ideal springs [117]. Reaction times between successive target acquisitions vanish in such a repetitive movement. This type of task, however, is not a very realistic scenario.

The meaningful manipulation of real or virtual objects requires the initial identification of the next target. Many experiments thus involve randomly changing target positions. Among the first, Fitts himself together with Peterson showed that his model also applies to such discrete target acquisitions, although with slightly different performance characteristics [88]. In comparison to the earlier study on cyclic aimed movements [87], they observed faster active movements, but an additional reaction time before movement onset.

In principle, the measured throughput of various direct manipulation interfaces could be compared across different studies. This should be done with great caution, since the many external factors that cannot be specified or controlled, imply large variations. Direct comparisons across different studies are further complicated by a lack of reliable data. Research publications on interface performance do not always report the empirically determined parameters of the regression model and often even do not provide the necessary data for computing the index of difficulty. However, a multitude of available studies on target acquisition allow the estimation of through-

---

[4] The target width divided by two represents the error tolerance for aiming at the center of the target.

[5] Throughput specifies the amount of information that can be transmitted per second.

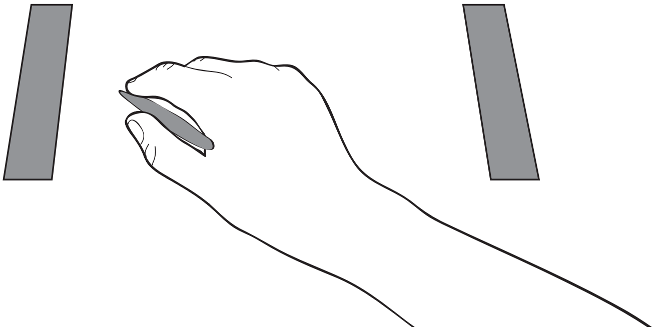


Figure 2.1: *Illustration of Fitts's aimed movement task. Participants were asked to move as fast as possible between two targets on a desk and hit them within specified boundaries. Target distance and width were varied [87].*

put as the quotient of mean ID with mean movement time. Note that this is not a constant value representing the input system performance, since the intercept a has a varying impact on the result of the division [377]. However, the effect of a only lowers the computed value. The resulting performance values are thus conservative estimates.

Further sources of error must be kept in mind when comparing target acquisition performance using Fitts's Law across different settings and conditions. For example, Graham and MacKenzie found that target width has a larger impact than distance and that this effect is even more pronounced for computer interfaces [109]. Therefore, the results of several studies cannot be compared formally, but they indicate a reasonable range of performance in fundamental target acquisition tasks. Results from various target acquisition studies can be critically reviewed with respect to such ballpark figures.

Even if the actual range of tested IDs is not reported, rough comparisons seem to be possible. The logarithmic term makes extreme IDs unlikely if reasonable target sizes and distances are applied. For example, reaching a 2 centimeter wide target area in 6 centimeter distance, corresponds to an ID of 2 bits, while the ID of aiming over 1 meter distance at a target of 2 millimeter width is slightly below 10 bits. Fitts's law generally does not hold for extremely small targets [58] and extremely large distances [56,119] (unless scaling is involved [118]) that can be more accurately modeled as a sequence of target acquisition tasks. Researchers generally avoid such extreme cases, if they are not deliberately interested in their effects, hence IDs beyond 10 are not to be expected. Commonly used IDs are in the range of 4-7. Therefore, task completion times that exceed the average in similar tasks by a factor larger than two, indicate that the task was extremely difficult or the tested interface does not support effective operation.

Computer interfaces can be tweaked to make target acquisition tasks much easier than they are in the real world. Snapping techniques effectively stretch the target widths in motor space [25, 36, 37]. Motion constraints reduce the dimensional complexity of target acquisition [35,335]. Non-linear transfer functions from user input to the manipulated virtual objects can shorten the target distance and increase its size in motor space. Pointer acceleration, for example adjusts the control-display gain based on the input motion velocity. Slow motion is further slowed down for higher accuracy and fast motion input is accelerated for faster target approximation (see [55] for a series of studies on this subject).

Such augmentation techniques are useful, but they should be considered a secondary layer of the interface beyond direct user input. Essentially, they trade the accurate representation of user input for efficiency on target acquisition. If the users' intentions are known, this can be an advantage, but such implicit interpretation of user input also reduces its potential expressivity[6]. Therefore, the benchmark for direct manipulation interfaces is real-world performance in corresponding actions. Also, user performance with interfaces that build on linear transfer functions can be considered as a reference.

2.1 2D-Pointing Benchmarks

Graphical user interfaces following the WIMP[7] paradigm, represent all application elements on the 2D screen surface and allow their direct access by pointing at the respective locations. This ever repeated action is an almost perfect resemblance of Fitts's classical pointing studies. Consequently, an evaluation scheme based on Fitts's law is included in ISO 9241 on the ergonomics of computer interfaces [91] and many reference results can be found in the literature.

Fitts's original studies on pointing with a stylus at a table revealed movement times between 180 and 781 milliseconds for IDs between 1 and 7 bits [87]. He measured slightly longer movement times for a 16 times heavier stylus (1 lb. vs. 1 oz.). In a pin-transfer task with IDs ranging from 3 to 10 bits, participants achieved task completion times between 326 and 959 milliseconds. Moving plastic washers between pins, took the participants between 535 and 1096 milliseconds for IDs from 4 to 10 bits. Fitts computed throughputs for the different conditions in the range of 10 to 11 bits/s with outliers for very small IDs. MacKenzie later computed the throughput based

---

[6] Freehand sketches, for example, contain much more subtle meaning than the sum of the represented objects and environments. They allow for ambiguity on aspects that are not yet fully defined, while key concepts can be highlighted. Construction plans, on the other hand, provide the required accuracy for the implementation of an idea. Computers are generally appreciated for their support on accuracy, but several drawing and sculpting applications also apply the unfiltered creative expressions of their users. Ideally, computer-aided design applications capture the original traces of user input and then derive accurate results from iterative specifications.

[7] Windows, Icons, Menu, Pointer

on the effective index of difficulty. For both tasks in which the target location was physically constrained, he confirmed Fitts's results of throughput in the range of 11 bits/s. The stylus conditions, however, where participants could miss the accurate target location, resulted in lower throughputs of 8.1 bits (1 lb.) and 9.2 bits (1 oz.) bits/s.

As mentioned above, Fitts's original experiment involved repetitive acquisitions of the same two targets. A later study on the acquisition of targets at randomly changing locations involved an additional cognitive load to identify the next target location [88]. Fitts and Peterson measured this reaction time separately from the active movement time [88]. Interestingly, the pure movement times in this study were 100–200 ms (30-50%) shorter for comparable IDs. Computing the throughput only from movement times thus yields an impressive level of about 16 bits/s. A possible reason for this advantage could be the effect of effort optimization during the repetitive task (c.f. [117]). If we take into account the almost constant reaction time of about 300 ms, however, the overall performance in these discrete target acquisition tasks drops to approximately 8 bits/s. These pointing tasks with a stylus are similar to icon selection with graphical computer interfaces and may serve as a real-world performance benchmark. Based on these experimental results, we can expect that target acquisition times with adequate computer interfaces should not exceed far beyond one second.

2D pointing input captured with mice, touchscreens, or digitizing pens generally enable task performance with a throughput of 4 to 6 bits/s (e.g. [92, 214, 215, 298, 366, 367]). Graham and MacKenzie directly compared virtual and real pointing with one's finger under very comparable conditions. From the their description of the study we can derive average throughput rates of about 9 bits/s for the real condition and 8.5 bits/s for the virtual condition [109]. Apparently, pointing can be performed on a comparable level in both cases. Several researchers that measured throughput as the reciprocal of the slope in the linear regression model, reported values between 10 and 12, which also corresponds to real-world performance [94, 261]. However, direct comparisons as well as those between average values of multiple studies consistently indicate slightly lower performance. The interaction with a graphical computer environment seems to involve a constant overhead compared to the real world. Computer interfaces offer less sensory feedback and they can be impaired by the effects of latency [217, 359] and tracking noise [260], which may impair the pointing performance.

The efficiency of pointing in graphical user interfaces might still allow for slight improvement. Alternatively, the information capacity of user interfaces can be extended with additional degrees of freedom. Our motor capabilities are not reduced to pointing at interactive surfaces. We can use our whole bodies in concert with sophisticated tools to perform complex navigation and manipulation tasks. Emerging interaction styles based on multitouch input, 3D user interfaces, tangible devices, and full body motion capture can take advantage of our real-world skills to enable more powerful and engaging computer applications.

2.2 The Information Capacity of Full Body Motion

A recent publication by Oulasvirta et al. offers an estimate of the input expressiveness that could potentially be exploited for interaction [253]. They suggested a novel method for measuring the information capacity of human motion input. Traditional throughput analysis of user input based on Fitts's law considers only the width and distance of a target as the conveyed information with the spatial deviation of actual hits as a measure of noise. Oulasvirta et al. 's method, instead, takes the whole motion trajectory, its shape and its velocity profile into account. They analyzed recorded motion sequences for their inherent entropy or information complexity. Pairwise comparisons with repetitions of the same sequence revealed a measure of reproducibility, hence, the amount of information in a sequence that was actually controlled.

They analyzed three different motor actions of varying complexity, with a common ISO 9241 mouse pointing tasks as the simplest case. For the task that generally results in throughput rates of 4-6 bits/s they measured about 37 bits/s with their alternative method. The whole trajectory of a movement, e.g. handwriting, obviously involves much more information than tapping a start- and an endpoint (Figure 2.2). Oulasvirta et al. also tracked motion sequences of a ballet dancer with 37 3D markers covering the whole body and obtained information throughput rates of up to 584 bits/s.

This is certainly an impressive leap in information capacity, but it is only useful for human-computer interaction, if the machine can derive meaningful interpretations from the transmitted information, e.g., if the path towards an object selection could define the desired mode of operation and the setting of involved parameters. Eventually, high data transmission rates are not a feature for its own sake, but we apply input to achieve certain goals. The common focus on the goal of an action is thus still a reasonable premise. The results of Oulasvirta et al., on the other hand, provide a good estimate of just how much more expressiveness could potentially be achieved by one user alone. Collaboration of many users can potentially multiply these expressive capabilities if the coordination overhead is small.

The large throughput measures of Oulasvirta et al. result from the huge amount of degrees of freedom that can be controlled with the body. For their measurements they used 37 3D markers that capture 111 degrees of freedom during motion performances that may serve as gestural input to computer applications. Following the paradigm of direct manipulation, instead, additional degrees of freedom can also be realized through interactive objects with extended manipulation possibilities. Instead of pointing at 2D screens, we could move application elements through 3D

Figure 2.2: *The trajectory of a movement between two points, can obviously involve much more information than it is expressed in the target distance and width, but handwriting, for example, is not an aimed movement.*

space while rotating them and changing their size and proportions at the same time. We could do so with several elements at once and thus engage with multiple hands or users in parallel.

However, the coordinated manipulation of several freely movable objects can be challenging, in particular, if multiple users are involved. Also the implementation of such interfaces is more difficult. System designers must consider more complex processes and concurrent manipulations of multiple application states. The potentially available degrees of freedom are thus often sacrificed for the sake of simplicity. Mobile handsets, for example, are mostly operated with only one finger although they are equipped with multitouch screens.

In an attempt to better exploit our human motor skills HCI-researchers developed a variety of experimental input devices and interaction techniques and analyzed the variables affecting the performance of their operation. The experiments were conducted with different tasks and under varying conditions. Thus, their results are not directly comparable, but as mentioned earlier, the different tasks can generally be broken down into simultaneous and subsequent combinations of aimed movement tasks in 2D or 3D space. Fortunately, Fitts's law seems to be applicable across different types of aimed movements. The general model has been shown to apply for 3D pointing[8] [57,113,145,241], object rotation (if the rotation axis is known or clearly visible) [230,326], and two-finger input like pinch zooming [338]. We can thus compare the results of different studies to benchmark performance in corresponding tasks under real-world conditions or with particularly efficient computer interfaces.

---

[8] Note that extensions to the model have been suggested for 3D pointing to account for comparably small effects of the motion direction.

2.3 3D-Pointing Benchmarks

3D pointing can be considered a small extension to its 1D and 2D counterparts. It seems to involve the same cognitive processes and very similar motor actions, but with one additional degree of freedom of the target location. Under real-world conditions, this does not seem to make much difference. The motion trajectories between tapping actions involve all three dimensions anyway and if the targets are within reach, also the number of involved limbs remains the same. A physical target object generally provides support that effectively prevents overshooting in one direction – independent of whether the targets are arranged in one, two, or three dimensions.

Our motor skills, however, seem to vary with the direction of movement. Apparently, target locations right in front of us can be reached faster than those that involve movement to the side or upwards. Empirical studies on these effects in real-world 3D pointing experiments have revealed differences in the range of 100 ms for pointing tasks with IDs between 2.5 and 6 [57, 241]. Murata and Iwase reported an average throughput of about 4.7 bits/s [241]. Perhaps, better performance would have been possible with a less encumbering measurement apparatus than the wired 3D tracking device they attached to the participants' fingertips. From the data provided by Cha and Myung on a similar 3D target acquisition task we can derive an average throughput of about 6 bits/s, which seems to better reflect our 3D pointing capabilities in the real world [57].

Nieuwenhuizen et al. compared the effects of physical or virtual targets on 3D pointing [246] using a wired 3D tracking stylus. From their task descriptions and the figures in the paper, we can derive target distances between 10 and 20 cm and a target width of approximately 1.5 cm, which corresponds to a maximum ID of about 4.3 bits. They recorded average task completion times of one second in the real-world condition (Figure 2.3). This seems relatively long for the short distances. A throughput of approximately 4.3 bits/s is at the lower end of real-world pointing performance. It can be speculated that the results were negatively affected by the apparatus and the test procedure. The target platform used in the experiments consisted of wooden cylinders at different heights. All cylinders were always present and before starting the movement the participants had to identify the respective target cylinder based on a dedicated number. This cognitive mapping task was certainly time consuming, although it was trained beforehand. Moreover, all other cylinders besides the current target posed obstacles to the pointing movement. Without such adverse conditions, it can be expected that 3D pointing performance in real-world settings is closer to the throughput of about 6 bits/s as obtained by Cha and Myung [57].

In the virtual condition of Nieuwenhuizen et al.'s study, the task took participants more than 50% longer, even though the targets were directly highlighted and could thus be recognized without effort. A detailed analysis of movement phases, revealed that the final correction phase for accurate pointer placement required most of the

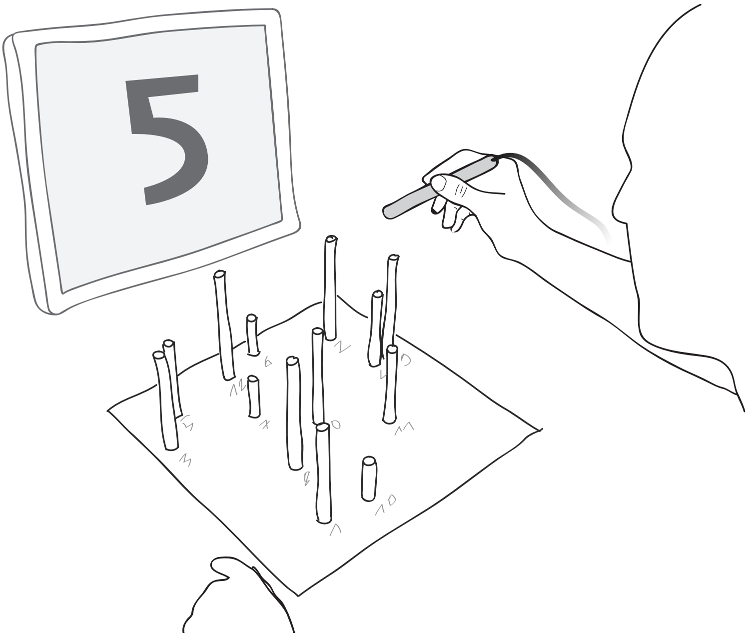


Figure 2.3: *Real-world 3D pointing setup of Nieuwenhuizen et al. The participants were asked to tap with a wired 3D pointer on wooden cylinders of different heights. The cylinders were randomly numbered and the target numbers for each trial were given on a computer display. The illustration is based on the task description and figures provided in [246].*

additional effort. Coarse target approximation during the ballistic phase took only about 15% longer. The participants were approximating the targets about 70% slower, but on a shorter path (since no physical obstacles were involved). The slower movement in this phase could be an effect of latency. The accurate target acquisition at the end of the task took more than twice as long as in the real-world condition. Movement velocity was about the same, but the motion paths were longer. This is most probably an effect of missing haptic feedback, which made participants overshoot the targets several times before selection. In the real-world condition, instead, a collision with the target ensured the correct final position.

3D pointing in virtual environments was studied extensively and generally revealed similar performance results as those of Nieuwenhuizen with throughput rates in the range of 2 to 2.5 bits/s [9, 113, 334] or lower [359]. The performance of 3D pointing at virtual targets seems to suffer primarily from latency [359] and a lack of tactile feedback, which reduces the achievable accuracy [246, 334]. Therefore, if low latency can be provided, large virtual targets may be approximated as fast as real-world objects, but accurate pointing will always require much more effort.

Despite the lower throughput rates, 3D target acquisition with corresponding 3D input devices can be more efficient than the same operation with 2D input devices like the mouse. With the latter the 3D space cannot be crossed diagonally and multiple manipulation steps must be performed. At minimum, two steps are necessary, e.g., to adjust the position on the x/y plane and then in depth. Considering a task with equal distance ratios between all three axes, we can assume an overhead in the index of difficulty of about 75%. Therefore, it requires approximately two times higher throughput rates for 2D input techniques to have the potential for more effective operation than integral 3D motion input. Switching modes between the different input mappings, however, adds another overhead that additionally involves cognitive load. Performance comparisons between 2D and 3D input for 3D target acquisition, thus depend on the throughput that can be achieved with both devices and the efficiency of the applied mode-switching technique.

2.4 3D-Rotation Benchmarks

3D object rotation can be very difficult. In principle, the optimal transition between any two orientations is a single turn that never exceeds 180°. However, the axis of rotation is not intuitively clear if it does not correspond to any of the principal axes of the object or the environment (Figure 2.4). Parsons showed that in this case people have a hard time to tell whether two objects in different orientations are of the same shape or to describe the shortest motion path between both postures in terms of rotation angle and axis [257]. If the target orientation cannot be identified before the motion onset, the applied rotation is not an aimed movement, but rather a search process. Consequently, coarse target approximation in a ballistic phase cannot be expected in this case, but rather slow and continuous motion that permits closed-loop control based on visual feedback. The movement time of object rotations would then exhibit a rather linear relation to the ratio of target distance and size.

Active object rotation alleviates the difficulty of mental rotations. By trying we can rapidly align two objects according to their geometrical features and compare their morphological similarity. Zhai and Milgram showed that any 3D rotation task can be successfully solved in interactive settings, although with lower efficiency in case of odd rotation axes [378]. Apparently, the achievable performance in rotation tasks depends on many more aspects than the size and distance of the target orientation. The geometric complexity of the rotation object and the orientation of the rotation axis seem to be particularly relevant. Fitts's law has only been shown to apply in cases where the angle and axis of the rotation are obvious [230, 326]. In comparisons of rotation performance in different tasks and settings we must expect a large variance. Nevertheless, best-case results from the literature may serve as benchmarks.

Ware and Rose compared 3D rotations of real and virtual objects [360]. Their results demonstrated that rotating real objects about 125° on average with a tolerance of 4-5°

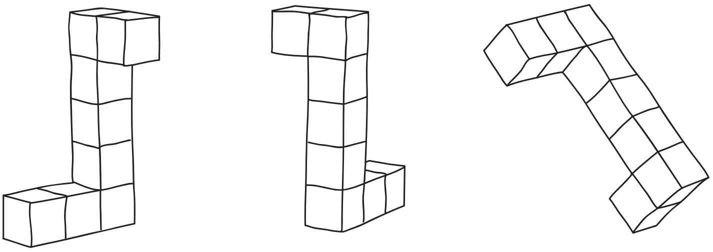


Figure 2.4: *Mental rotation tasks with a Shepard-Metzler object (see [257]). The comparison of two objects in different postures and the specification their transformation is particularly difficult if the rotation axis does not correspond to a cardinal axis of the object or the viewer. The transformation between the figure shown in the center and that on the right is such an example. The rotation axes between the figure on the left and both others corresponds to a cardinal axis of the object and is therefore more comprehensible.*

can be achieved in less than two seconds [360]. This corresponds to a throughput of about 2.6 bits/s. In a blindfolded condition it took the participants about 17% longer to achieve the target rotation with an average error of about 9.21°. If the same task was performed with the same physical handles as input, but visual feedback from computer graphics, participants could achieve a similar accuracy as in the real-world condition, but it took them 25% longer. Ware and Rose suggested that the reason for the slightly lower performance was the system latency of about 75 milliseconds. Other studies on 3D rotation of virtual objects generally obtained even worse results than those reported by Ware and Rose (e.g. [138,202,274,358,379]).

2.5 Object Manipulation Benchmarks

So far we have analyzed the potential user performance in turning knobs and pushing buttons. Moving objects in space involves simultaneous 3D rotation and translation with six degrees of freedom (DOF). User performance in such 6-DOF object manipulations is generally studied with docking tasks that require the 3D alignment of two equally shaped objects. It is open to debate, whether the simultaneous control of rotation and translation is efficient and whether 6-DOF manipulation can be modelled as an aimed movement task that follows Fitts's Law. In any case, turning objects during displacement is a common behavior with mundane objects. Consider, for example, moving a telephone handset to the ear in order to take a call. Such simultaneity of rotation and translation has also been observed during the manipulation of virtual objects with various 2D and 3D input devices [190,326,356,357,381].

Masliah and Milgram, on the other hand, demonstrated that such simultaneity is not necessarily efficient [222]. Their results indicate that performance benefits cannot be expected from simultaneous rotation and translation. The index of difficulty in 6-DOF manipulation tasks might thus be adequately defined as the sum of two aimed movements that describe both submovements separately. In fact, Stoelen and Akin, found a good fit for this simple model, although the participants of their study operated rotation and translation simultaneously (all but one of 13) [326]. For the sake of comparability between aimed movement tasks with different numbers of DOF, we use this combined index of difficulty to estimate performance in the following discussion of experimental results on combined rotation and translation. Note that this measure is a very conservative estimate that neglects potential benefits from simultaneous rotation and translation.

The results of Stoelen and Akin on user performance in their 2-DOF docking task[9] may serve as a reference for simple object docking in computer applications. For combined IDs in the range of 5 to 12 bits, they obtained task completion times between 2 and 3.5 seconds which corresponds to a mean throughput of about 4 bits/s. This seems quite low for such a simple task. Wang et al. earlier studied almost the same task with a wooden cube on a physical table [355,356]. They tested object docking with translation distances of 30, 100, and 200 mm and rotation offsets of 22,5° and 45°. They also reported docking accuracy with mean translation overshooting of 2.6 mm, mean rotation undershooting of 2.5° and variable errors of 1.9 mm for the translation and about 2.5° for the rotation subtask [355]. This data allows the computation of an effective index of difficulty of 7.4 bits. The obtained average task completion times of 776 milliseconds thus correspond to an effective throughput of 9.6 bits/s. This performance is more than two times better than that obtained by Stoelen and Akin for virtual object manipulation but it compares quite well to the performance in real-world pointing.

Experiments on docking virtual objects with more involved degrees of freedom generally have revealed much worse performance. Ware reported 14.05 seconds for a 6-DOF docking task with isotonic position control [357]. We can estimate a combined ID of about 10 bits (using the average error for target width), hence the index of performance would be in the range of 0.7 bits/s. Zhai and Milgram showed that users can strongly improve 3D object docking performance with training and reduce the task completion times from almost 20 seconds to less than 7 seconds[10] [378]. These results were later confirmed by Masliah and Milgram using almost the same task and the same input device [222].

The task description in both papers does not directly support the computation of a combined ID since target distances were not reported explicitly and the required accuracy was given in terms of distance tolerances of the four vertices of an equilateral tetrahedron. Considering the description of the apparatus and the required accu-

---

[9] translation and rotation in only one dimension each

[10] including 0.7 seconds dwell time at the target as a task completion criterion

racy[11], maximum combined IDs of 8 to 10 bits can be estimated. The results would thus correspond to a maximum throughput of about 1.5 bits/s.

Zhai and Milgram also showed that position control is generally faster in such tasks than rate control while the latter allows for higher accuracy and more efficient motion trajectories. Kunert et al. tested 6-DOF object docking with an isotonic 6DOF wand device while standing in front of a large 3D powerwall and obtained average task completion times of 4.1 seconds for relatively coarse placement with an average combined ID of about 6 bits [190]. Thus, a similar throughput as in the experiments of Zhai and Milgram could be achieved. They also confirmed the earlier observed differences between position control and rate control.

2.6 Conclusion

The above reviewed studies on user performance in elementary manipulation tasks offer reference values for the comparative evaluation of novel interaction designs. The derived throughput estimates do not satisfy the requirements for statistical comparison or scientific proof, but, they serve as ballpark figures that indicate the potentially achievable performance. Comparisons between different studies must further take into account that the effects of many additional factors are not considered in the computation of throughput. This is particularly true for comparisons between different tasks since more complex tasks are not simply the sum of elementary actions.

However, we can observe some general trends in the data. While the throughput in 2D pointing tasks with computer interfaces can get close to real-world performance, this is by far not the case for 3D pointing or 2D docking tasks. Real-world performance seems to be less affected by geometrical task complexity. Only for 3D rotations the throughput is comparably low in real and virtual conditions. In both referenced conditions of Ware and Rose's studies [360], the participants rotated objects in mid-air without physical support. Therefore, physical constraints could be one of the main factors for target acquisition performance. This could also explain the large differences between real and virtual conditions in 3D pointing and 2D docking experiments.

The target locations in most 3D virtual environments are only visually indicated [9, 113,246,359]. In real-world experiments on 3D pointing, instead, the target is a physical object that prevents overshooting at least in one dimension [57,241]. In the real-

---

[11] 8.4 mm tolerance for each vertex of a tetrahedron with a side length of 4.2 cm allows maximum rotation errors of about 18.5° – if the center is placed perfectly accurate. In practice, an angular accuracy of about 9° seems to be achievable [360] and still retains half of the given positioning tolerance, hence these values were used for the estimation of the indices of difficulty. Assuming a 20 inch CRT display as in [222] the available interaction space allows maximum translation distances of about 20 cm from a corner to the target center. Rotation distances are limited to 180° at maximum.

world 2D docking task studied by Wang et al., the target location was not fully constrained and thus overshooting could occur [356], but in comparison with the study of Stoelen et al. on combined rotation and translation with a handheld tracker [326], the test participants could benefit from a desk providing physical support for the manipulated object. Several studies showed such benefits of constraints in direct manipulation tasks (e.g. [176,296,334,335]).

Further aspects should be considered as potential factors of the apparent differences. The slightly higher 2D pointing performance in real-world environments – even compared to touch devices and digital pens [94, 298] – indicates a general disadvantage of computer interfaces. The performance difference could be attributed to the fact that Fitts's original studies involved only target variations along a single dimension, while computer interfaces are often tested with actual 2D pointing tasks as suggested in ISO 9241 [91], but several of the reported evaluations of computer interfaces actually used 1D tasks as in Fitts's original studies (e.g. [214, 298]). The performance difference between pointing at 1D and 2D targets is fairly small [145], if the height of the target in movement direction is not much smaller than its corresponding width [3, 114, 213]. Tracking jitter and system latency must be considered as factors [166, 260], but with state-of-the-art 2D interfaces this should be a minor issue. Perhaps, we operate computer interfaces slightly different than purely mechanical setups. When hitting a physical target location we immediately consider the task completed, while we expect an additional feedback from computer applications. This expectation alone might account for slightly longer task completion times with computer interfaces.

Seemingly, the information capacity of pointing-based 2D user interfaces cannot be much further improved. More expressive user input to direct manipulation interfaces can thus rather be realized with higher input complexity. Unfortunately, more complex operations seem to take over-proportional time, when performed with computer interfaces. This is true for 3D pointing and rotation alone. Their combination in 3D docking tasks yields even lower performance. In relation to the combined IDs of the involved rotation and translation tasks, performance measurements hardly exceed a throughput of 1.5 bits/s.

The simultaneous operation of multiple degrees of freedom thus appears to be detrimental to the task performance in virtual object manipulation. If this overhead exceeds that of separating the task into subsequent 2D pointing tasks, the potential benefits of simultaneous operation cannot be exploited. As mentioned above, the distance overhead resulting from DOF separation in 3D pointing, increases the index of difficulty by maximally 75%. A similar overhead is involved when operating 3D rotation with screen-space techniques [141]. However, if the costs of mode switching can be neglected, mouse- or touch-based interfaces may be the better choice since they enable faster target acquisition [31, 335]. The following chapter reviews literature on cognitive processing and motor capabilities to gain a better understanding of the parameters governing the integrality and separability of degrees of freedom in direct object manipulation.

Chapter 3

Integrality and Separability

Jacob et al. compared the control structure of input devices with the perceptual structure of task attributes in terms of their integral and separable dimensions [157]. They demonstrated superior user performance if both were corresponding. For example, object motion across 3D space appears to be better supported with input devices that provide integral control of the three degrees of freedom. Dissociated parameters, instead, such as brightness and 2D position, can be adjusted more efficiently with separate control devices.

The theory was derived from findings in experimental psychology on the perception of object differences which showed that similarity ratings either reflect the sum of differences on multiple unrelated dimensions or, instead, the shortest distance across a coherent multi-dimensional attribute space [102, pp. 98–102]. Garner explained this observation with perceptual differences between integral and separable attribute dimensions and argued that "... dimensional structure is important for separable dimensions and similarity structure is important for integral dimensions." [102, p.111]. Objects that differ in separate attribute dimensions (e.g. shape and color) are sorted according to the dimension with more salient differences, e.g. all objects of various shapes but the same color would be separated from all objects of another color. In case of integral attributes like brightness and color saturation, instead, classifications are based on perceived distances along one common similarity scale, i.e., a diagonal across all involved dimensions [102].

The theory of integral and separable object attributes relates to semantic differences. Some attributes and their manipulation are more interrelated than others and it seems obvious that attributes of color should not be confused with attributes of shape. Interface designers would intuitively provide separate control facilities for both – even without knowing anything about the research on perceptual integrality and separability. A concept of related and unrelated attributes does not require much understanding of cognitive psychology. Garner and his colleagues, however, were interested in understanding human attention and recognition processes and observed that integral and separable attributes lead to different classification processes. Integral dimensions cannot be distinguished in primary processing. The perceived similarity of objects is thus corresponding to the Euclidian distance in a coherent attribute space. Moreover, attribute integrality hinders the classification of stimuli based on a single dimension. Integral dimensions are generally perceived in combination. If, instead, objects differ in separable dimensions, we tend to consider only the most salient one in classification tasks (see [102]).

3.1 Attribute Interrelations in Perceptual Processing

The usability of user interfaces benefits from a control structure that reflects the integrality or separability of the manipulated attributes [157]. Unfortunately, we cannot build on either a comprehensive taxonomy or a conclusive definition of integral and separable dimensions. Garner and colleagues developed methods for the identification of integral attributes in classification tasks, but they also observed contradictory results. Experiments on the comparison of rectangles with varying width and height, for example, revealed indications of dimensional integrality [102, p.140] [84] but also separability [102, p. 165] [361]. The differences can be explained with the perceptual strategy of stimulus redefinition: combined changes along integral or separable dimensions can be interpreted (redefined) as another, potentially more discriminable, dimension [102, p. 133]. Following this concept, width and height can be perceived integrally in terms of corner-point distance or surface area, but they separately affect the nominal dimension of form, which is easier to distinguish than changes of width and height.

Assessing the integrality or separability of attribute dimensions is not obvious. Garner contemplated that integrality is a continuum rather than a dichotomy [102, p.127] and argued that only a limiting definition can be provided on the basis of logical relations between the involved dimensions. He suggested that integral dimensions are mutually dependent, i.e., "... in order for one dimension to exist, the other must be specified ..." [102, p.136]). For the separable case, instead, he argued, that "All combinations of existence and nonexistence of dimensions are possible." Following this definition, only few attribute pairs can be considered as truly integral. We cannot represent hue without involving values of saturation and brightness. These three dimensions of the Munsell color scheme[1] are thus considered to be integral. This does not imply, however, that the three attributes could only be manipulated in combination.

---

[1] http://munsell.com/

Width and height of a rectangle also seem to fall into the category of integral dimensions. We cannot specify one without the existence of the other if the shape is meant to remain a rectangle. Moreover, the ratio of both defines the specific shape of the rectangle, which seems to be a more salient perceptual attribute. The position of a point in 2D space is a less obvious case. The interrelation of vertical and horizontal position is generally considered a prime example of integrality, but is it a relation of existential dependency? Arguably, the context largely affects the amalgamation of dimensions. We have a clear concept of a one-dimensional position, but it cannot exist solely in a two-dimensional context. Similarly, vertical and horizontal position fully define a location in 2D space, but in a 3D space they must involve a value of depth.

In their experiments on the correspondence of task attributes and control structure, Jacob et al. considered size and 2D position as three integral dimensions of a manipulated object [157], although this interpretation is not directly supported by the referenced research in experimental psychology. Garner and colleagues generally compared only two attributes at once. One of the few exceptions was a study on dimensional preferences for the classification of two-dot patterns with three varying attributes: horizontal position, distance, and orientation. Imai and Garner observed that single attributes were used for sorting a set of stimuli that varied equally over all three dimensions. Different participants of the study chose different attributes for their classification, but generally a single one was selected. Imai and Garner assessed these clear preferences for one of the three dimensions as a proof of their separability [102, p.134] [153]. Under the assumption that distance and size are closely related, these results indicate separability rather than integrality of position and size.

One could argue, that size and position of an object are nevertheless integral attributes of a rectangle. They must both be defined in order to perceive an object somewhere, but the same is true for the relation of two dots. Based on such a loose interpretation of Garner's definition, the specification of a rectangle also requires some color and other appearance attributes in order to distinguish it from the background. Intuitively, the size of a rectangle appears more closely related to its position than its color. Position and size are both geometric attributes, while color is not. The study of Imai and Garner, however, demonstrated that geometric attributes can be clearly separable too.

The integral processing of separable attributes requires some sort of stimulus redefinition or a perceptual change of context. The size of an object, for example, is directly related to visual distance. Size in combination with vertical and lateral object positions thus indicates a 3D position. The mapping of input motion to the manipulated attributes in the study of Jacob et al. fostered such a mental model. They reported that "Movement of the wand in a plane parallel to the screen moved the user-controllable object in x and y. Moving the wand toward the screen made the object either bigger or darker; away made it smaller or lighter" [157, p. 11]. The combined manipulation of size and 2D position thus resembled 3D object movement in front of a mirror – a mental model that maps the separate task attributes into a joint manipulation space

(Figure 3.1). The simultaneous manipulation of position and brightness did not support such stimulus redefinition.

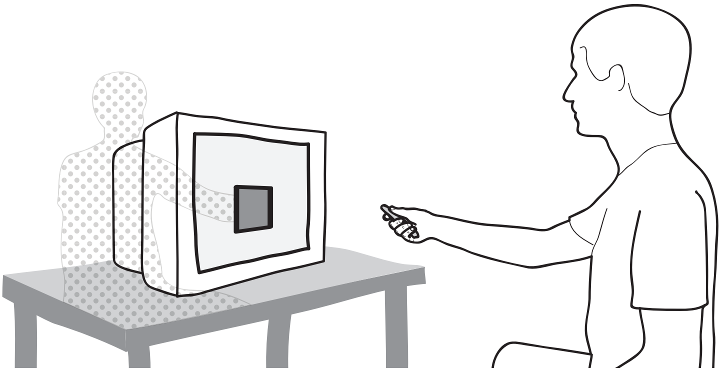


Figure 3.1: *Illustration of Jacob et al.'s test setup [157] for the integral manipulation of position and size of a rectangle. Lateral and vertical input movements with a handheld 3D tracked stylus was mapped to the corresponding motion of a rectangle on the screen. Input movement in depth changed its size. Moving the stylus closer to the screen increased the size of the rectangle and vice versa. This mapping corresponds to moving an object in front of a mirror. The gray dotted figure in the background shows the mirrored pose of the test user to illustrate this mental model. No such mirrored user representation was visible during the studies.*

The theory of stimulus-response compatibility suggests that input mappings must be compatible with the operator's mental model for efficient interaction [275, pp. 287-290]. It implies a context for the perceptual processing of the involved dimensions. Therefore, the input mapping could be more relevant for the perceived integrality of the involved DOF than the inherent integrality or separability of the task attributes. If the input mapping supports the representation of the involved degrees of freedom in a joint attribute space, they can be processed and operated integrally.

The interrelations of attributes also depend on perceptual focus. Garner addressed this issue with the concept of asymmetric integrality of attributes [102, p. 136]. Some attributes can be treated separately if they are the center of attention, but they cannot be ignored, in turn, during the processing of other attributes. As an example, Garner refers to an experiment in linguistics on the discrimination of spoken syllables. The two involved factors were start consonant and pitch [102, p.137]. Day and Wood found effects of pitch on the discrimination of the phoneme, which indicates an integral perceptual structure of both attributes [74]. The phoneme, on the other hand, had no effect on the discrimination of the pitch value, which indicates separate processing of both factors. The observation follows Garner's general definition of integrality. A spoken consonant or vowel must involve some level of pitch in order to exist, while pitch exists independently of being used for the articulation of a phoneme, hence they affect each other's processing asymmetrically. Garner suggests the asymmetric integrality of shape with attributes of color and size as another example. While the perception of shape interacts with both other dimensions, color and size can be processed separately from shape [102, p. 137].

According to Garner, the integrality of dimensions is an inherent, and therefore mandatory, property of a stimulus, while separability denotes the option for separate processing of dimensions. The human brain can apply alternative processing strategies to facilitate classifications that involve separable attributes [102, pp. 132-135]. Selective serial processing, for example, builds on the primary selection of the more discriminable dimension for subsequent classification. Stimulus redefinition, another perceptual strategy that was mentioned above, combines separable dimensions to a new one that is easier to distinguish. Garner also notes that the integrality of attributes does not negate their perceptual distinctiveness. Otherwise, the successful Munsell color scheme, which consists of only integral dimensions, would be nonsense. The difference between integral and separable attributes is dominant in primary perceptual processes as occurring in speeded classification tasks, whereas the inherent dimensional structure can be overruled in secondary perceptual processes. Ashby et al. referred to perceptual vs. decisional integrality to describe the effects of different perceptual stages [11] and emphasized contextual effects on the integral or separate perception of attribute dimensions.

Several experimental results have indicated varying levels of attribute integrality, task dependency and integral processing of separable dimensions. Garner, therefore, proposed the idea of a perceptual continuum between the integrality and separability of attribute dimensions [102, p. 127]. Other researchers supported this idea (see [273] for an overview). Hyman and Well, for example, suggested a "continuum of combining rules" that affect the perceptual interrelations of attribute dimensions [152].

In an experiment on effects of spatial relation on attribute processing, Potts et al. presented the attributes of size and orientation as the size of a circle and the orientation of a radius line stretching from the center to the contour. Although size and orientation are generally considered to be separable dimensions, they observed asymmetric integral processing in this configuration of stimuli [273]. Spatial proximity of stimuli with separable attributes seems to foster their integral processing. Other experiments confirmed the effects of spatial relations of attributes on their perceptual interaction. Integral attributes like brightness and saturation, for example, were shown to be processed separately when perceived through dissociated objects or positions [103]. Shepp suggested that this effect of spatial configuration can be accounted for by extending the continuum of dimensional interaction from integrality over separability to (spatially) separate [315]. Similar effects on the perceptual interaction of inherently separate attributes can be expected from other types of perceptual grouping.

3.2 Integrality and Separability in Manipulation Tasks

Jacob et al. suggested that knowledge about the perceptual processing of dimensional structures in classification tasks could directly be applied to the design of manipulation interfaces [157]. Their pioneering work inspired many researchers to design and evaluate novel user interfaces according to the notion of separable or integral degrees of freedom (e.g. [21,139,159,216,278,345]). The above review of research in the field of experimental psychology, however, casts doubts on the reliability of the concept as a design guideline. Perceptual integrality and separability may be inherent to the attributes of a task, but this does not necessarily imply how they can be manipulated most effectively. The concept is not a dichotomy and it does not define static relation between dimensions. Instead, our perception in classification tasks is biased by several external variables that include the context, the perceptual focus and the spatial relation of stimuli. In more interactive settings, perhaps, the processing of dimensional structure also varies with the type of manipulation task and interaction method.

Research on similarity perception and object classification, has considered a variety of stimuli that include color attributes, acoustic phenomena and also spatial dimensions. We build on highly trained and partly even innate mental operations for the classification of perceived phenomena. However, in the context of manipulation interfaces, the captured user input can be incompatible with the operated dimensions. Research on stimulus-response compatibility shows how non-corresponding mappings impede human interaction performance (for an overview see [275]). It has also been shown, that certain non-isomorphous mappings are more compatible than others [275, pp. 171-193] and that several dimensions appear to be perceptually overlapping [182], but in general, stimulus and response should be corresponding with each other and to the mental model of the human operator.

Most user interfaces are based on spatial motion input. They enable movement of virtual objects and symbolic input via keystrokes or touch events at specific locations. Spatial manipulations can thus be supported directly with compatible mappings, but the adaptation of other attributes requires mediating controllers, e.g., a color selection widget, that translates 2D motion input in screen space to color attributes. The appearance and behavior of the controller serves as a cognitive bridge from the available input space to the target attributes.

In one condition of Jacob et al.'s experiments, three-dimensional motion input was directly mapped to two different attribute spaces, namely brightness and 2D position. The integral operation in such different attribute spaces requires a common mental interaction model that offers a compatible mapping for both. A coherent interaction model can even facilitate the integral manipulation of separable object attributes, as we have seen in the joint manipulation of 2D position and size. In the case of manipulating position and brightness through 3D motion input, one might also consider various mental models that could map all involved dimensions to a joint manipulation space, e.g. moving an object through a foggy 3D environment. Moving the object back and forth would affect its brightness as a function of distance to the viewer. This far-fetched model, however, would also imply corresponding changes of the object's size.

Applications of the theory on integral and separable attribute dimensions in the field of interaction design, have usually been concerned with spatial manipulation. Researchers questioned which spatial degrees of freedom could be considered integral and which are rather separable. Earlier research on information processing indicated that position, orientation and distance are perceived and processed separately [153], while variations in vertical and horizontal position are processed integrally [103]. If these results can be generalized to account for spatial manipulation, they suggests that different types of spatial manipulation are separable, while the three Cartesian dimensions should be treated integrally.

Wang et al., proposed a different interpretation of Garner's findings. They argued that "..., object transportation and orientation could be integrable because the spatial attributes are generally considered integral" [356]. As discussed above, this does not seem to be the case. Garner and colleagues did not consider all spatial attributes to be integral. They even demonstrated the separability of position, orientation and distance [153].

Wang et al. made an important point, however, when they emphasized the cognitive differences between the perception of passive objects and active manipulations. They referred to research results indicating separate visual processing systems for perception and action (see [108]), which question the applicability of Garner et al.'s observations to the design of manipulation interfaces. The position of an object could be perceived very differently than its active translation and the perception of an object's orientation could differ strongly from the perceptual processes involved in its rotation. They further referred to research on human motor control suggesting independent visuomotor processing of attributes that relate to different actions or phases of actions like reaching or grasping an object (see [259]). Assuming that object rotation and translation may affect manipulation planning and motor control differently, they reasoned that both might also be processed and performed separately.

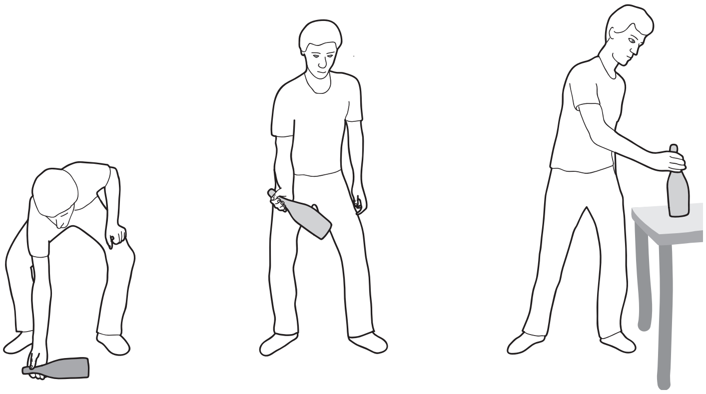
Wang et al. analyzed the simultaneity and the interdependence of translation and rotation during object placement [356]. Their docking task required the alignment of a small wooden cube with a corresponding virtual wireframe model. The results show that translation and rotation were operated concurrently, but every process started and ended with translation-only phases. Rotation thus appeared to be a sub-operation of translation. The interdependence of both processes was demonstrated by mutual interference with individual task performance. A larger target distance of one attribute also increased the operative time on the other, but translation had a stronger effect on rotation than the other way around. The reduced performance on the individual tasks indicates a coordination overhead for the combined manipulation.

The coordination overhead does not necessarily hinder effective operation. We computed an effective throughput of 9.6 bits/s for object placement with three degrees of freedom in the study of Wang et al. (see Section 2.5). This performance is quite similar to Fitts's measurements in a one-dimensional tapping task (see Section 2.1). The similar level of throughput in both cases indicates that the combined task also took correspondingly longer, hence, the participants could not induce more controlled input in the same time – as one might expect from simultaneous operation. Operating both subtasks subsequently, might have been just as effective, if mode switching would be effortless. However, switching modes from rotation to translation or vice versa generally interrupts the manipulation process and adds cognitive load.

Masliah and Milgram proposed the m-metric for measuring multi-DOF coordination as the product of simultaneous control and task efficiency [222]. The m-metric only considers simultaneity of successful error reduction, not simultaneous action as such. Simultaneous error reduction on multiple axes is furthermore only recorded as the minimum, instead of the average value on all involved dimensions. This makes the m-metric particularly sensitive to reduced efficiency on individual dimensions that may result from simultaneous operations. The metric was used to analyze two devices that afford integral 3D rotation and translation: an isometric rate controller and an isotonic position controller. All six degrees of freedom were operated simultaneously with both devices, but the integral operation was most effective for 3D translation and 3D rotation alone.

This is not surprising, as the simultaneous 3D manipulation in a joint attribute space (position or orientation) enables diagonal shortcuts, whereas the combination of 3D rotation and translation, may not reduce the target distance. Improved efficiency can still be expected from parallel execution, but only if, the increased coordination effort does not outweigh the benefits. Using the m-metric Masliah and Milgram observed a reduction of efficiency by about 40% if both were operated simultaneously. These measurements provide a deeper understanding of the observed interdependence of rotation and translation in the study of Wang et al. Simultaneous operation reduced the individual efficiency and thus an increased effort in one operation also caused decreased performance on the others.

Several studies have shown evidence that humans have a tendency for the simultaneous execution of rotation and translation if this is not inhibited by constraints of the manipulated object [222, 326, 356, 357]. Numerous everyday examples show such movement integration. For example, when moving a glass to our mouth for drinking, aiming a key towards a door lock, or picking up a bottle from the floor, we can observe simultaneous rotation and translation in many routine object manipulations (Figure 3.2). Presumably, this behavior is not or only partly related to the goal of improving time efficiency. Human motor control is not directly operating on the degrees of freedom of objects in the environment, but it must coordinate the multitude of degrees of freedom of our limbs, which imply other motion constraints that can be beneficial or detrimental to various manipulation tasks.



(a) Underhand grip (b) Rotation & translation (c) Placement on table

Figure 3.2: *Real-world object manipulation often involves simultaneous rotation and translation. Human motor-control is optimized for the movement capabilities of our body and the physical constraints of our environment – not so much for time efficiency and an optimal trajectory in world coordinates. The optimization is apparent also in unconscious action planning. We tend to pick up objects in a more uncomfortable hand posture, e.g. an underhand grip (a), to provide for a comfortable end state after an object rotation (c).*

3.2.1 Implications from Human Motor Control

The geometric relation of an object with the surrounding environment is only one aspect that affects movement planning and coordination during its manipulation. Like all human movements, object manipulation requires control over multiple limbs, each with several degrees of freedom and particular constraints. As a result, many different movements can be performed in order to achieve the same goal. This freedom is certainly an advantage as it facilitates the avoidance of obstacles, however, it also complicates control. The neurophysiologist Nikolai Bernstein highlighted the “degrees of freedom problem” in human motor control [33] and inspired many subsequent researchers to explore this issue. One plausible solution is the cost containment theory (see [290] for an introduction). The theory states that human motor control aims to to minimize various costs related to postures at the start, the end, and during a movement. Rosenbaum et al. suggested that depending on the situation different weights are assigned to the costs that can be relevant to the required motor performance [290]. These weights allow the actor to put an emphasis on accuracy, speed, collision avoidance, style, or any other factor that may be considered relevant.

Rosenbaum et al. proposed internal representations of posture as a basic building block of human motor planning, because these allow simpler internal representations than trajectories [290, p.178]. This view is supported by studies demonstrating that humans have difficulties in memorizing movement, while postures can be easily recalled. Moreover, postures can be specified in terms of equilibrium points for the muscles, i.e., "a set of muscle lengths for which muscle tensions balance out". Rosenbaum et al. argue, that "when an equilibrium point is specified and the starting point is known, the trajectory to the equilibrium point comes for free, making detailed planning of the trajectory unnecessary." They also refer to the "end-state comfort effect" which describes a behavioral tendency to grasp objects in an uncomfortable posture in order to achieve a comfortable posture at the end of the manipulation process. A bottle lying on the floor, for example, will generally be taken with an underhand grip in order to facilitate its rotation to an upright orientation (Figure 3.2 a). Therefore, simultaneous rotation and translation of objects may be inefficient in terms of motion trajectories in world coordinates, but it may be the result of efficient limb coarticulation (see also [286, p.22]). If computer interfaces do not enable similar handling of virtual objects, our motor planning and operative skills may be impaired.

The central nervous system seems to exploit various strategies to simplify effective motor control (see [34] for an overview). The choice of the most suitable frames of references for motion planning and operation is one example. Berthoz mentioned egocentric and allocentric (related to external coordinates) reference systems. He emphasized the role of gravity as a reliable natural reference and argued that our brain exploits reference frames connected to the limbs in order to simplify motion control. The action of pointing towards a remote target, for example, can be simplified to the control of two polar coordinates centered in our shoulder joint [34, p.107]. Moving an object with a similar strategy implicitly involves its rotation relative to an external coordinate system, while it remains stable in relation to the operating hand and arm. If further variables such as the distance or the orientation of an object in the hand should be kept stable, kinematic constraints provide implicit coordination of the limbs. One of these internal control loops is the movement of connected limbs in phasic opposition to satisfy a certain motion constraint, e.g. drawing a straight line: "When the angle of the arm increases in relation to the body, the angle of the arm in relation to the forearm decreases by an equal amount" [34, p.144]. These insights indicate that human motor control may facilitate simultaneous rotation and translation as a byproduct of maintaining relations among the involved limbs.

The manipulation of objects is not only a mechanical task, but it requires cognitive effort for planning the action and monitoring its execution. This can be particularly difficult for 3D rotation. Translational movement of objects along the shortest path across 3D space can be readily imagined. Understanding the most efficient 3D rotation path, instead, has been proven to be difficult [257]. The same studies of Parsons also showed that the mental process is considerably simplified if the rotation axis coincides with one of the principal axes of the egocentric reference frame or the object's shape (see mental rotation tasks in Figure 2.4). Note that the handling of everyday objects primarily involves rotations of this type – and above all rotations about the vertical axis of gravity. When actively handling objects instead of only thinking about it, the task seems to be alleviated by continuous visual and proprioceptive feedback of the manual process and the continuously changing state of the object (see for example [378]). Mundane objects furthermore have constraints embedded in their shape and their distribution of mass. We exploit gravity to let objects swing into new orientations and let them align through collisions with planar surfaces like tabletops. Handling real-world objects is an iterative learning process that eventually enables playful interactions as in balancing, throwing and juggling.

3.2.2 Integral Attribute Manipulation with Separable Input

Intuitively, 3D object transport seems to be most compatible with integral 3D translation input. However, other mappings may also be suitable. Balakrishnan et al., for example, proposed design adaptations to the common mouse to better support 3D interaction [21]. The underside of their Rockin' Mouse was rounded such that the device afforded tilting. Tilting about the depth axis was mapped to translation along the vertical axis, while the common two-axis translation of the mouse was used to control virtual object motion on a ground plane. This design enabled the simultaneous operation of translation across 3D space – although with two different types of movement. They demonstrated that 3D positioning could be performed up to 40% faster with the Rockin' Mouse compared to a common 2D mouse with mode switching for additional degrees of freedom. An analysis of motion paths revealed that 49% of the motion input with the Rockin' Mouse involved all three dimensions simultaneously. Simultaneous translations along the three axes occurred primarily in the ballistic phase, in which users achieve coarse approximation. During the subsequent closed-loop phase, in which more fine-grained adjustments were performed, a tendency for axis separation was observed.

This example is another demonstration that rotation and translation can be operated simultaneously – at least at the motor level. An unusual mapping enabled the integral operation of 3D translation without lifting the arm. This approach seems to compromise stimulus-response compatibility, but fortunately, our mental models are flexible and can adopt to changing configurations. Research on stimulus-response compatibility shows that some non-isomorphous mappings are more comprehensible than others [275, Chapter 9]. Balakrishnan et al. mapped clockwise tilting of the mouse to upwards motion, while counterclockwise tilting moved the cursor down. With the mouse located to the right of the display, this mapping corresponds to Warrick's principle [275, Chapter 9.2]: It describes the perceptional preference for mappings from rotation input to translation where the linear motion direction corresponds to the closest tangent of the rotary controller. Corresponding examples are known from rotary knobs for the tuning of old-fashioned radio receivers. In case of the Rockin' Mouse (in right-hand operation), the left side of the device is closest to the display and adequately moves up and down during rotation (Figure 3.3). We can thus perceive a coherent mapping from 3D position input to 3D position control if we consider the position of the left side of the device as the input coordinate.

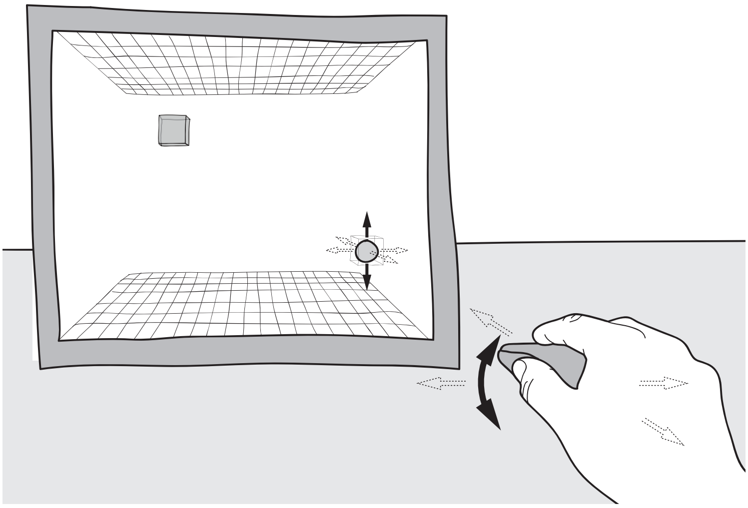


Figure 3.3: *The Rockin' Mouse of Balakrishnan et al. [21] enables full 3D translation with the device comfortably resting on the desktop surface. Tilting the mouse moves a 3D cursor (sphere with wireframe bounding box) up and down. The applied mapping corresponds to Warrick's principle [275, Chapter 9.2]. The experimental task of Balakrishnan et al.'s study was to place the 3D cursor inside a one third larger translucent target cube (in the upper left corner of the screen in this illustration).*

The question remains, whether this input mapping enables efficient manipulation of the three degrees of freedom. Balakrishnan et al. compared 3D object positioning with the device to a mouse-based technique and reported a performance advantage of approximately 30%. Not much detail on the task was provided, but from the description of the setup with a visual interaction volume of approximately 30 cm^2, we can derive a maximum index of difficulty of about 5. They recorded mean task completion times of 5.5 seconds for this task after training over five blocks of task repetitions. This would correspond to a very low thoughput of less than 1 bit/s. The report is not clear whether the task required spatial selection of the manipulation object in the Rockin' Mouse condition. If this was the case, then the task would have consisted of two aimed movements in 3D space with a combined index of difficulty of 10 bits at maximum. The throughput would then correspond to approximately 1.8 bits/s, which is still slightly less than the average throughput rate for 3D pointing tasks with unconstrained 3D motion input (see Chapter 2). The Rockin' Mouse does not seem to support the known performance of 2D pointing with the mouse in 3D pointing tasks.

In the 2D mouse condition, participants had to select one of three visible faces of a cube and then move it along the corresponding geometric plane. Therefore, in the mouse condition, the task consisted of four subsequent aimed movements, each with two degrees of freedom. Moreover, since the bounding box of the manipulation object was aligned with the screen plane, the perspectively distorted faces that enabled motion in depth offered only a small target width. The combined index of difficulty of the four subtasks in the mouse condition can be estimated to be in the range of 15-16 bits. Considering average throughput rates in 2D pointing tasks (about 4.5 bits/s), movement times should not exceed 3-4 seconds. Balakrishnan et al. recorded average task completion times of about 7.5 seconds after training. Apparently, the switching from one subtask to the next was also time consuming. The example of the Rockin' Mouse, therefore, demonstrates that the integral manipulation of degrees of freedom can be beneficial, despite the required coordination overhead. The elimination of mode switching allows users to perform integral movements in a coherent action. In this particular experiment, however, the mouse could have performed much better if the motion direction was toggled with mode keys.

Another notable example of 3D placement with separable input is the Balloon selection technique presented by Benko and Feiner [29]. They suggested to use multitouch input for the placement of a 3D cursor following the metaphor of a floating balloon on a cord (Figure 3.4). The x/y position is controlled with one hand moving along a touch surface. The distance between two fingers of that hand defines the size of the cursor. Movement perpendicular to the touch surface can be controlled with another hand inducing further touch input relative to the first one. They compared this technique to direct 3D pointing with a 3D wand and a keyboard technique where movement along all three axes was manipulated completely separate with discrete keystrokes.

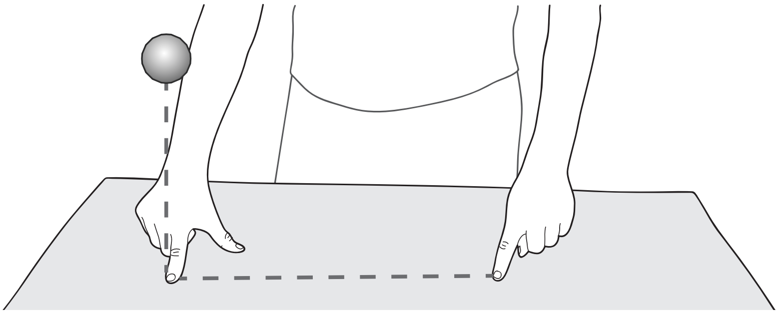


Figure 3.4: *Benko and Feiner suggested Balloon Selection. A multitouch input technique to control the position and size of a 3D cursor. One hand controls the x/y position of a 3D cursor above the touch surface. The distance between two fingers of that hand defines the size of the cursor. Additional touch input by another hand controls the height of the cursor by adjusting the distance between both hands.*

The task description, allows us to estimate an index of difficulty in the range of 3-5 bits. The average task completion times for the 3D wand and the Balloon technique were in the range of 6-6.5 seconds, while it took about 12 seconds on average with the keyboard technique. Considering a typical error rate of about 4% the throughput of the two faster techniques would be in the range of 0.75 bits/s. This is not necessarily a convincing performance, but in fact, only in the keyboard condition, users achieved an error rate of 4.1%. With the balloon technique it was slightly higher at 5.5%, but with the 3D wand, the error rate increased to 16.1%. This means that the effective throughput with the 3D wand technique was much lower than it was with the multitouch technique.

One reason for the overall performance below average may be the limited perceptual quality of the output device used in the study. The stimuli were presented on an optical see-through head-mounted display with a resolution of 800 x 600 pixel. A more plausible reason, however, seems to be the required accuracy. For target volumes of 10, 8, and 6 mm^3, the authors measured average error rates across devices above 6%. For the smallest target size of 4 mm^3 it went up to about 14%, which indicates, that Fitts's Law might not hold any more, because the target size was below the physically achievable accuracy. Moreover, the experiments involved additional adjustments of the cursor size, which added variability on yet another separate dimension.

In any case, the comparison of the Balloon technique with integral 3D motion input indicates a competitive edge of separable over integral input for 3D aimed movements. The reason for this advantage is perhaps the physical support provided by the multitouch sensor, which facilitates accurate placements. The distribution of control between both hands also seems to be advantageous in this case. The cooperative bimanual 3D motion control supports simultaneous as well as separate adjustments of the involved degrees of freedom. In the following section we will discuss further examples of input combinations with multitouch interfaces.

3.2.3 Integrality and Separability in Multitouch Interfaces

Multitouch interfaces have introduced a considerable increase of input capabilities compared to earlier established technologies such as the mouse. Moreover, the combination with spatially congruent visual output, as realized with touchscreen devices, has motivated the design and implementation of novel interaction techniques that mimic direct interaction with physical matter. It is an implicit paradigm of touchscreen interaction, to maintain the congruency of fingertips on the screen and the initial contact point with the displayed graphics. In case of at least two fingers touching the same 2D graphics, this implies the integral operation of translation, rotation, and scaling. Therefore, the availability of multitouch input devices has facilitated the integral control of multiple object parameters and revived the debate regarding integral and separable degrees of freedom.

Nacenta et al. argued that the available degrees of freedom should remain separable in order to avoid involuntary manipulations [242]. They proposed five different multitouch input techniques that allowed both: separate control of translation, rotation and scale for higher accuracy as well as their simultaneous operation for coarse target approximation. A formal study on object docking performance revealed significant benefits of this approach. They also studied the characteristics of object manipulation with unconstrained multitouch input. Measures of gestural noise, i.e. the magnitude of involuntary input, showed that object orientation could be kept close to the initial state during translations, while large magnitudes of involuntary scaling occurred. A temporal analysis of the largest motion magnitudes revealed the concurrent operation of all three manipulation types during the ballistic phase. Corresponding to the findings of Wang et al. [356]), they also observed rotation as a sub-operation of translation.

Nacenta et al. also applied Masliah and Milgram's m-metric [222] to the data which revealed relatively high measures of efficient simultaneous operation in particular for the combination of rotation and translation. The obtained value of 0.43 corresponds to the level of simultaneous control Masliah and Milgram recorded for pairs of translation input [222]. The m-metrics for translation and rotation in combination with scaling both reached a value of 0.32. This indicates that simultaneous operation was less effective than the combination of rotation and translation, but still partially useful. We can conclude from these results that the simultaneous operation of translation, rotation, and scaling with multitouch input can be efficient, but it may cause erroneous input, which in turn can be minimized with sophisticated interaction techniques that support voluntary input separation.

Multitouch input can also be applied to 3D object manipulation. Several mapping techniques have been proposed, most of which build on the established two-finger input for combined rotation, translation, and scaling [126,131,181,206,221]. The latter is often applied to movement in depth instead of scaling and additional fingers or single-finger input are used for the remaining two rotational degrees of freedom. Reisman et al. suggested a more implicit technique based on a constraint solver that dynamically computes 3D manipulations which satisfy the changing position of the contact points with the manipulated object in screen space. In comparative studies, however, more explicit mappings resulted in higher manipulation performance [206, 221].

Regarding the implications of perceptual integrality and separability of 3D manipulations with multitouch input, different observations have been reported in the literature, sometimes contradicting the intuition of the authors. Martinet et al. agreed with Wang et al.'s interpretation that orientation and translation were integral aspects of 3D manipulation tasks [221]. This interpretation would support the design of integrated control interfaces for 3D rotation and translation with six degrees of freedom (DOF). They found advantages, however, of explicit separation between rotation and translation as compared to other DOF distributions that do not relate to these separate attribute spaces. On the contrary, Herrlich et al. observed performance benefits of tighter control integration in 6-DOF manipulation with multitouch input. The participants of their study, however, noted that DOF separation could support higher accuracy [131]. Veit et al. argued that rotational degrees of freedom should better be controlled separately because the difficulty of mental rotation tasks [257] inhibits optimal path planning across all dimensions. Their experiments, however, failed to show clear benefits of separate 3D rotation input [347].

3.3 Conclusion

Jacob et al. suggested that the control structure of input devices should reflect the perceptual structure of the tasks they are meant to be used for [157]. They referred to research on the perceptual processing of object attributes for the classification of integral and separable dimensions. The corresponding research results from cognitive psychology, however, do not offer an unambiguous taxonomy of integral and separable object attributes. Instead, it has been shown that the perceptual structure of object attributes depends on external factors like the context and the perceptual focus. Moreover, the perceptional structure of object attributes cannot be directly applied to their manipulation. The mapping from user input to the manipulation of various attributes also affects their perceptual structure. Research on stimulus-response compatibility offers a robust basis for the analysis of these interrelations [275].

The integrality and separability of object and task attributes probably affects motion planning and closed-loop motion control. If for example, an untrained manipulation task consists of a difficult rotation (e.g. with a skewed rotation axis) and a complex translation path (e.g. to avoid obstacles), we will take advantage of the separability between both attributes and process the respective motion requirements individually. For simpler tasks like placing objects on a desk or well-trained ones, like eating with fork and knife, we may instead combine both actions in motion planning and execution. Note that user interfaces with unusual operational principles, e.g. based on rate control or with abstract mappings from 2D input to 3D object manipulation, do not support the application of such real-world motor skills and thus generally require more training.

The perceptual structure of object attributes does not imply whether the concurrent manipulation of these attributes is efficient or not. The human locomotor system applies other constraints than those of external objects and environments, which favor simultaneous operation in many cases. More specifically, our motor control seems to use varying reference systems to simplify motion planning, many of which are related to skeletal joints rather than objects and the environment [34]. This strategy is particularly effective during open-loop ballistic movement, which is involved in aimed movement tasks for coarse target approximation. For accurate placement, instead, we must consider external reference systems. Consequently, a tendency for concurrent or integral manipulation of many degrees of freedom can be observed during the ballistic phase of movements [131, 242, 356, 357]. Towards the end of an aimed movement, when accuracy becomes key, advantages of DOF separation have been observed instead [21, 251, 348].

User interfaces should ideally support both, simultaneous and individual operation of multiple degrees of freedom. Such adaptability can be achieved with filtering methods and advanced transfer functions (e.g. [97, 242, 348]). Alternatively, interaction techniques can build on the combination of multiple inputs to extend or reduce the number of simultaneously available degrees of freedom. Multitouch systems, for example, allow users to extend single-touch 2D position control with simultaneous rotation and scaling through the involvement of further touch inputs. Modifier keys, on the other hand, allow users of layout and design software to constrain 2D motion input to a single axis. Chapter 5 reviews prior research on such combinations of multiple simultaneous input facilities.

Chapter 4

Cooperative Action and Movement Coordination

Cooperative action can increase the efficiency, expressivity and fluency of various operations and performances. Clay modelling, playing musical instruments, and even everyday activities like walking are impressive examples of a highly coordinated cooperation between multiple limbs. Groups of people in orchestras and dance companies show that such a tight interweaving of actions can also be achieved among multiple people. The required coordination is certainly more challenging in the latter case, but most people have well developed capabilities for interpersonal cooperation – at least to support each other in mundane actions like the joint handling of mundane objects (Figure 4.1). Cooperation extends the available degrees of freedom which in turn enables more complex actions or simply more flexibility to cope with dynamically changing situations.

Cooperative actions may occur simultaneously or sequentially. They can be further distinguished to be symmetric or asymmetric. Simultaneous actions in the same direction increase the power of a movement, e.g. to pull a heavy object. If applied in different directions, it facilitates control of distance or size (e.g. squeezing and stretching). In asymmetric cooperation, on the other hand, actions build on each other as in a chain of hinged levers. This in turn enables actions of higher complexity.

Symmetric actions are often performed simultaneously, as in the case of grasping and lifting a large object with two hands. Asymmetric actions, on the other hand, generally involve subsequent steps that build upon each other. In the previously mentioned example of clay modeling, we can observe both types of cooperative action in fluent alternation. The modeler uses both hands in asymmetric operations, if one hand moves the clay and holds it in position to support the other hand’s subse-

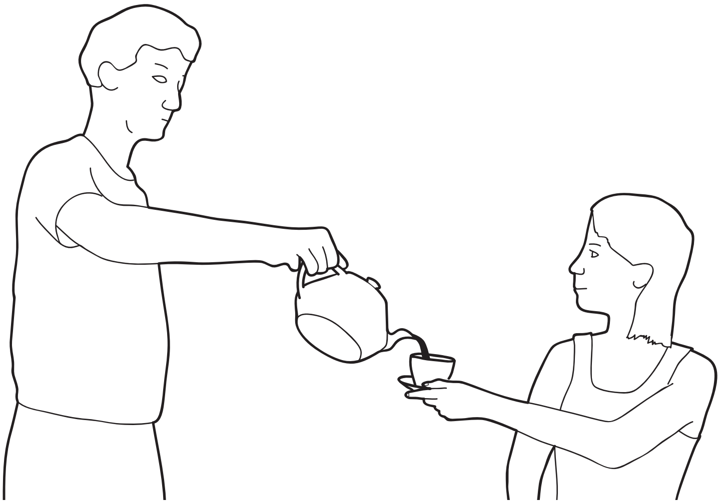


Figure 4.1: *Pouring tea into a cup that is held by someone else is essentially a distributed aimed movement.*

quent creation of dents and bumps. Squeezing the clay, instead, involves two fingers or hands symmetrically and simultaneously. Many other movement patterns of multiple involved limbs or actors involve symmetric actions that occur sequentially. This is the case in walking, climbing, crawling, and similar alternating movements of multiple limbs that extend the immediately available motion range of a single movement. Passing loads in a human chain, e.g. a bucket brigade, is a similar example involving multiple people.

4.1 Synergies

Cooperative action requires coordination. This is not only an issue for joint action of multiple people or the cooperation of left and right limbs, but the whole locomotor system builds on the tight coordination of interconnected limbs, each with several degrees of freedom. Reaching for an object may involve any number of them. Although the targeted object position can be readily described with a three-dimensional vector, our motor systems involves at least seven degrees of freedom that define the relation between hand and shoulder. As a result, there are several options for how to reach the object, and even more so if we consider contributions of further body parts. This flexibility is beneficial as it supports the adaptation to changing situations, but it poses a coordination problem to our central nervous system. Almost any goal of a motion can be achieved in many ways that involve different contribution of our limbs.

Bernstein first identified this degrees of freedom problem. He contemplated that muscles must be co-operated to reduce the coordination effort, hence, a single high-level command must activate the movement of related limbs which implicitly complement each other's contributions. He also realized, on the other hand, that a one-to-one correspondence of voluntary motor impulses and resulting motion trajectories cannot exist. The high-level command cannot be hard wired to a specific contributions of the involved limbs [33]. During the course of any movement the situation is continuously changing. Depending on the posture, the influence of gravity changes, weights and velocities affect further inertial forces, and potentially, moving obstacles pose varying counterforces to the planned movement. The central nervous system builds on feedback loops (often proprioceptive) to incorporate all these dynamic parameters for effective motor control. Bernstein argued that "the cerebral motor area organizes responses by deftly adjusting and balancing between resultant external forces and the manifestation of inertia, constantly reacting to proprioceptive signals and simultaneously integrating impulses from separate central subsystems, so that ten successive repetitions of the same movement demand ten successive impulses all different from each other; ...". [33, p. 33]

Observations of repeated aiming movements supported these hypotheses. Motion recordings of a blacksmith working with hammer and chisel, for example, revealed that the motion variability of involved joints and limbs exceeds that of the resulting effector trajectory (the trajectory of the hammer) (see [33, p. 69] and [197, p. 15], the original reference [32] is only available in Russian). Apparently, the movements of connected limbs compensate each other's errors. Bernstein described this behavior as synergies, which meant the grouping of muscles and limbs that can be controlled in conjunction. Through such synergies, a particular movement can be described, and potentially also controlled, with a smaller number of variables than the degrees of freedom that are actually involved. The resulting dimensional reduction can be mathematically evaluated through principal component analysis [68].

Latash later argued, that dimensional reduction does not solve the problem of redundancy and it cannot explain the flexibility with which humans perform similar tasks [198]. Together with Gelfand he reframed the problem of redundant degrees of freedom as a beneficial principle of abundance that enables multiple alternative motor actions to achieve the same goals [196, p. 202] [105]. An obvious example is the operation of a door handle with the elbow, while both hands are busy carrying bulky and/or fragile objects (Figure 4.2). It is precisely the complexity of our motion apparatus that allows us to perform effective motor actions under the most diverse conditions and to maintain stability of the most relevant parameters (e.g. the orientation of a cup filled with a hot beverage) even in case of external disturbances. Bernstein had observed this remarkable adaptability and identified the synergistic cooperation of the involved muscles and limbs as a key to efficient motor control. An accurate definition of synergies and their role in motor control, however, remained unclear. Several different configurations of multiple limbs can achieve the same goal in a synergistic way, hence the redundancy of options cannot be solved by dimensional reduction alone.

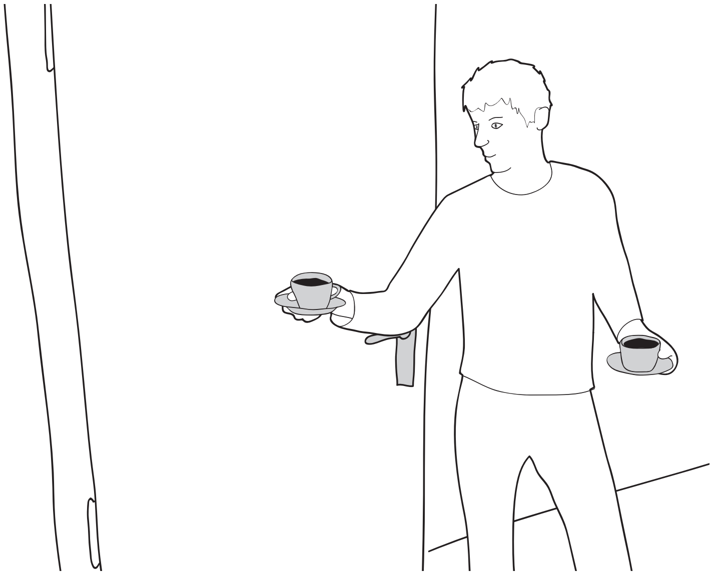


Figure 4.2: *Our motor apparatus incorporates many degrees of freedom that allow us to perform motor tasks in a variety of ways. At first sight, the abundance of possibilities may appear detrimental as it complicates motor control, but it supports dynamic adaptation to changing situations. For example, we can open a door with the elbow while carrying cups of hot coffee.*

Based on the above mentioned principle of abundance, Latash recently offered a more specific definition of synergy as "a neural organization that provides for low variability (high stability) of an important performance variable by co-varied adjustments of elemental variables" [198, p. 159]. The performance variable in this definition relates to the goal of a motor action. The elemental variables are the available degrees of freedom of the musculoskeletal system. According to this definition the potential configurations of elemental variables are reduced to a subspace in which the performance variable remains stable. If one or several of the elemental variables change as an effect of external disturbance or voluntarily in order to perform a secondary action (e.g. pushing the door handle), other degrees of freedom must co-vary in order to remain in the configuration subspace that leaves the performance variable unaffected. The subspace of goal-equivalent joint configurations can be mathematically described as an uncontrolled manifold (UCM) [306]. In a simplified case of only three degrees of freedom, this uncontrolled manifold would be an isosurface along which the performance variable remains stable. Elemental degrees of freedom may freely adapt to changing circumstances as long as they remain in this subspace, hence the name uncontrolled manifold. Deviations from this subspace, however, result in erroneous motor performance. The concept allows the quantification of synergistic coordination as the ratio between variance that occurs within the uncontrolled manifold (desirable flexibility) and that occurring orthogonal to this subspace (flawed performance) [199,306]. If the permitted parameter space described through the uncontrolled manifold is not otherwise constrained by the situation at hand, the central nervous system is considered to optimize system configurations and actions for minimized effort [290,337].

Synergistic covariance of limb movements implies their temporal synchronization. If the limbs, involved in a movement, try to compensate each other's errors in a post-hoc manner, the resulting movement trajectory will be unstable, hence their covariation must occur simultaneously. Indeed, synchronous covariation of multiple limbs can be observed in locomotion (see [33,287]) as well as pointing and grasping actions (see [288]), but such ingenious motor behavior is not a given. It requires tremendous training and mutual attunement of the involved actors.

Bernstein and colleagues conducted extensive research on motor coordination in human locomotion [33, p. 60]. They measured forces, velocities, and accelerations at the different limbs during walking and running. Their motion recordings with scanning frequencies in the range of 200 Hz allowed them to identify subtle details of the dynamic motion structures. They found highly complex motion substructures in all involved limbs during a single step and showed that these subsequent phases of varying acceleration and velocity are very similar across adult humans in general. These activity sequences constitute reactions to the motion of neighboring limbs in the kinematic chain. The observed similarity of these complex motion sequences thus reflects an optimal coordination of limb activities during walking with a human body. The rhythms and amplitudes of these sequences differ between people with different walking styles, but, the fundamental action sequences of interrelated limbs was found to be very regular. The complex dynamic structure is a result of a long learning process in childhood and its complexity vanishes in old age. A decrease of agility is also apparent in a simpler structure of motor sequences, with less synergies between involved limbs.

According to Bernstein three stages of learning and development can be distinguished. When learning to walk, small children at an age of about two years do not show much regularity of the dynamic phases and struggle with the external effects of gravity, collision and inertia. It generally takes until about five years of age before proprioceptive feedback becomes incorporated in the motion sequence and motor coordination improves. At this stage proprioceptive feedback is incorporated only after the pose resulting from the prior motor impulse can be perceived. As a consequence, the motion sequence is not yet very stable. At an age of about ten, feed-forward proprioception enables the simultaneous covariation of the involved limbs as described through the concept of synergies. It can be expected that the development of synergies in other motor actions also requires significant training.

4.2 Hierarchical Control

Many students of cognitive psychology subscribe to the concept of cognitive control hierarchies (see [33,210,364]. Models of hierarchical control explain how the invariant goals of our actions can be repeatedly achieved despite considerable variation in the contributions of multiple involved limbs which are orchestrated subconsciously. Skilled musicians, for example, perform highly complex motor actions while paying conscious attention on the resulting phrases and melodies instead of individual notes or finger movements.

Hierarchical models can also explain the observation of motor equivalence, i.e., the transfer of movement skills and styles from one motor subsystem to another. A common example is the formal consistency of one's handwriting if other limbs are used for moving the pen: writing with the feet or mouth results in highly similar traces to habitual handwriting with the dominant hand (see [195]).

Logan and Crump recently made another strong argument for the hierarchical model of cognitive processes. They showed how the various observations from studies of skilled typewriting can be explained with nested control loops that are largely independent and only communicate via a narrow interfaces. More specifically, they suggested an outer loop concerned with language processing that activates an inner loop controlling the execution of keystrokes [210].

4.3 Bilateral Coordination

The synergetic coupling of mechanically joined limbs such as the lower and the upper arm clearly simplifies motor control. A single motor command may suffice to realize motion towards a new posture of the whole kinematic chain. The coordination between the left and the right hand or both feet, appears more complicated at first sight. They are not directly connected, but only through the whole body. The states of several intermediate joints, each with multiple degrees of freedom, affect their spatial relations. Moreover, the left and the right side of the body are controlled by different brain hemispheres. In case of symmetric activities, both sides could be provided with closely related commands to assume spatially congruent or mirrored postures. However, such a parallel control structure would not enable synergistic behavior like reciprocal error compensation, demonstrated in bimanual pointing tasks [77,238].

Guiard demonstrated the predominance of asymmetric cooperation in mundane bimanual actions. He argued that both hands are coordinated according to a kinematic chain. Despite the lack of a direct physical link, both hands are considered to operate like two serially assembled motors [115]. Following this model, the non-dominant hand initiates a bimanual action by providing a spatial reference frame for successive actions of the dominant hand. An often quoted example of this behavior is the active involvement of both hands in writing. One hand holds and occasionally moves a sheet of paper while the other one is writing on it [115].

The kinematic-chain model redefines the concept of human handedness. Instead of a general preference for one of both hands it assumes the prevalence of bimanual interaction with asymmetric roles. Moreover, it embeds the actions of both hands in a longer chain of actions that includes the whole body. The non-dominant hand is considered to be specialized in postural support and initial reach, while the dominant one manipulates objects with higher force, rapidity, and precision. Hinckley et al. demonstrated increased accuracy and faster performance in a bimanual target acquisition task if the roles were distributed according to Guiard's model [138]. They also showed that the distribution of roles is less relevant if the task requires less accuracy and haptic guidance is provided.

The different roles are not inseparably tied to the left or right hand. Most people consider themselves to be either left- or right-handers, some assign the roles differently depending on the task, and ambidextrous people use both hands interchangeably. In principle, we all can switch the roles of our hands in everyday routines. Most of us will experience a lack of proficiency of both hands in the unfamiliar roles, but with additional effort, we can exchange roles even in asymmetric bilateral actions. When arriving at home with a heavy shopping bag in our stronger hand, for example, we may find ourselves opening the door lock with the non-dominant hand – although this is more difficult. Highly trained tasks or those that do not require high precision can be easily swapped between hands. Proficient tennis players, as another example, swap hands for a more balanced match against less trained opponents. In this sense, we may also consider antiphase symmetric actions like climbing and walking a special case of asymmetric cooperation with continuously alternating roles of both actors. While one foot is standing on the ground it provides a stable platform for a step of the other one. This step establishes the next intermediate point of support for the same action with reversed roles. If it comes to more difficult steps or those requiring more force (e.g. jumping), however, we generally revert to a clear preference of a dominant foot [258, 264].

4.3.1 Bilateral Movement Synchronization

Just as the synergies between directly connected limbs, also bilateral coordination implies temporal synchronization. Several studies have shown that it is difficult to perform repetitive tasks like finger tapping with both hands in different rhythms (e.g. [178, 263]). Instead, people tend to move their limbs synchronously, either in phase or antiphase. Coordinating simultaneous bilateral actions that are truly out of phase is difficult. Musicians train extensively for playing conflicting rhythms concurrently (i.e. polyrhythms, irrational rhythms). The complex performance can be facilitated by restructuring the rhythms and melodies in chunks and phrases that better relate with each other. The accentuations of a three-four time and a four-four time, for example, coincide every 12 quarter notes. Presumably, the beginning, ending, and the adaptation of trained movement sequences involves more cognitive load than its continuation and repetition. Evidence has been found to show that memorized motor programs can be performed without the intervention of sensory feedback [168], while switching tasks and task sequences involves a cognitive effort that becomes apparent in measures of rapidity and accuracy [304].

Besides implicit movement synchronization, also a tendency towards movements in phase at higher frequencies has been repeatedly shown (see [124, 169]). The participants in a study by Kelso, for example, were asked to move both hands in antiphase to the beat of a metronome [169]. When the frequency increased beyond the individual preference (generally between 1 and 2 Hz), the coordination of both hands fell apart for a few cycles and then stabilized again, but, with movements of both hands in phase. When asked to perform symmetric hand motions, the coordination remained stable across frequencies. One can reproduce these results easily with finger tapping on a table. Kelso argued that these observations could explain how animals automatically adapt their gait to changing speeds.

A number of similar experimental results have indicated that human motor behavior may be ruled by the principles of dynamic systems, rather than clearly specified cognitive programs. For coordinated movements of both hands, for example, mathematical models of coupled oscillators were suggested [124, 307, 371]. Haken et al. demonstrated that such a model also predicts the observed phase transitions from asymmetric to symmetric bimanual movements [124].

An earlier study by Kelso et al. showed that temporal synchronization also occurs in non-repetitive aimed movements [171]. The experiment required two independent aimed movement tasks to be performed with both hands in parallel, but at targets of different size and distance. Fitts's law predicts that the easier tasks will be performed faster, but instead, both hands performed in synchrony with almost identical movement times. Peak velocity and acceleration differed correspondingly to reach targets at different distances, but their profiles over time were almost identical. Kelso et al. suggested that the observed temporal covariation could be explained with shared afferent signals, i.e., shared control over both movements.

In a series of studies with humans and monkeys Wiesendanger et al. demonstrated that also asymmetric bimanual tasks are highly synchronized [364]. One experiment demonstrated how phasic synchronization facilitates posture stability in bi-manual unloading tasks. Corresponding to their observations, a waiter can hold a tray stable while unloading plates and glasses because changes in the muscle tensions of both hands are activated simultaneously (Figure 4.3). Another experiment required the opening of a drawer with one hand and the subsequent removal of an object with the other. For all human participants, motion onset and peak velocity of both hands were almost in phase with 56 ms delay on average, although the individual goals (drawer, object) were necessarily reached sequentially (about 300 ms difference on average). On the one hand, these observations demonstrate largely parallel execution and tight temporal synchronization of asymmetric bimanual actions. On the other hand, they also show that the initiation and major events (e.g. reaching the individual targets) of the different actions occur one after another. The time differences were small, but, the grasping hand always followed the opening hand. The same synchronization pattern was observed in monkeys.

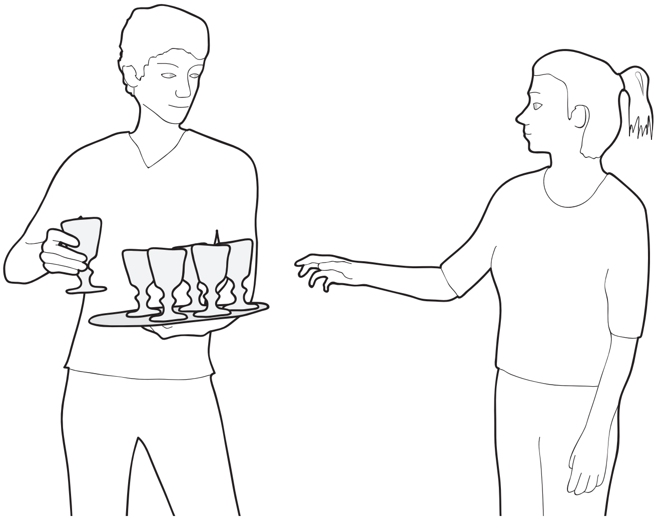


Figure 4.3: *Unloading glasses from a hand-held tray requires tight synchronization of both hands to maintain a stable posture despite the changing weight distribution. Wiesendanger et al. demonstrated this bimanual synchronization on the level of muscle activations as measured with an electromyograph (EMG) [364]. A waiter must achieve a similar quality of synchronization with the actions of another person to allow her taking a glass from the tray.*

4.3.2 Synchronization through Perception-Action Coupling

The tendency for the synchronization between different limbs has long been explained with models of shared afferent signals that co-activate linkages of muscles (e.g. [171]). More recently, it has been shown that the coordination, in particular bilateral coordination, may instead occur on the perceptual level (e.g. [96, 228]). Mechsner et al. showed that bilateral movement synchronization is more strongly influenced by visual perception, but also proprioceptive feedback may guide the coordination [228]. A particularly convincing observation came from an experiment in which participants operated the rotation of two visible flags (or levers) through cranks invisibly mounted under a table. A gear system applied a 4:3 transmission rate for one of both mechanisms (Figure 4.4). The participants were instructed to circle both visible flags either in phase or antiphase at the given rhythm of a metronome. While drawing circles at 4:3 frequency ratios is extremely difficult when directly observing the actions of both hands, the participants had no difficulties in performing this motor action with the manipulated visual feedback in this setting. Even a tendency of falling into in-phase rotation (of the two visible flags) as known from earlier studies [54, 169] was observed. If the coordination of both hands were governed solely on the basis of shared afferent signals, this behavior could not be explained. As an alternative, Mechsner et al. suggest that perception and action share a common cognitive representation which facilitates movement coordination through perceptual goals.

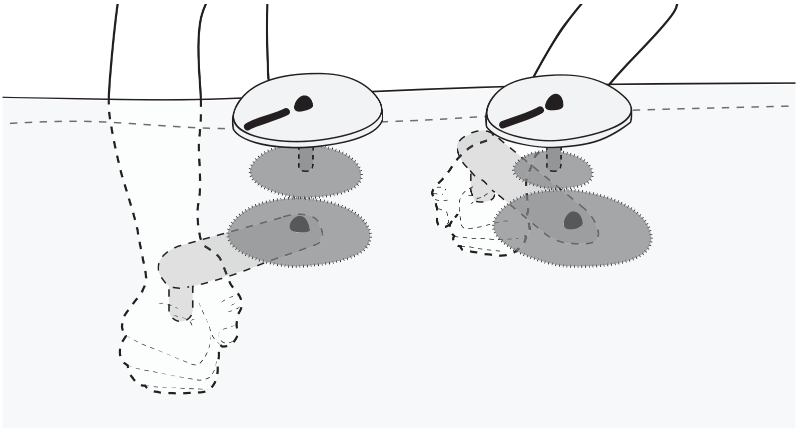


Figure 4.4: *Bilateral movement synchronization with different transfer functions. Mechsner et al. showed that turning cranks with both hands at different velocities (e.g. in a 4:3 ratio) can be performed without difficulties, if the resulting visual feedback compensates for the difference [228]. The participants of their study were asked to operate cranks that were mounted under a table. The movements were applied to visible flags (marked disks in this illustration) via gearboxes with different transmission ratios.*

Kelso et al. studied pattern formation in the coupling of perception and action using the example of synchronizing movements of a single finger to the beat of a metronome [170]. They found similar coordination dynamics as shown earlier for bilateral limb coordination and a systematic phase shift relative to the difference between the metronome frequency and the participants' preferred rate of finger flexions. As a consequence, they suggested an extended model incorporating the notion of the difference between requested and preferred frequency. The model predicts the synchronization of two systems (i.e. frequency locking) with a phase shift proportional to the difference of their eigenfrequencies. If the preferred frequency of the reacting entity (e.g. the moving finger) is smaller than the given one (e.g. the metronome beat), it will lag behind, and otherwise it will lead the sequence. If the difference is too large, synchronization will not occur.

4.4 Interpersonal Coordination

Further evidence for a perceptual foundation of movement coordination comes from studies on interpersonal cooperation. Schmidt et al. compared intrapersonal with interpersonal interlimb coordination in oscillatory movements (swinging pendulums). In both conditions, they observed strong rhythmic coupling effects and a preference of in-phase over antiphase oscillation [301]. Although the coupling strength was smaller in the interpersonal condition, these observations offered further evidence, that movement coordination can be governed by perception-action coupling and does not necessarily require shared motor control. Later research of Schmidt and colleagues confirmed the robustness of interpersonal synchronization, also called entrainment (see [303]). Apparently, the temporal coordination of independent actors can be predicted with Kelso et al.'s extended coupling model that incorporates effects of individual detuning from a preferred to a shared motion frequency [170].

Also, synergies can be observed in interpersonal cooperation. Riley et al. analyzed coordination in a target acquisition task with the operation of target and pointer assigned to different persons. Their evaluation revealed temporal synchronization, dimensional reduction and mutual error compensation as known from synergistic motor control [283].

Mottet et al. compared the performance in a one-dimensional, cyclical aimed movement task between four conditions: one-handed operation of either targets or pointer, simultaneous bimanual manipulation of pointer and target locations, and the distribution of control over pointer and target locations between two persons [238]. Their results demonstrate very similar behavior and performance in all conditions. Independent of the distribution of control, the recorded acquisition performance followed Fitts's Law with indices of performance in the range of 5-6 bits/s. In the conditions that allowed the simultaneous manipulation of target and pointer, both movements were highly synchronized, no matter if they were operated by one or two persons (Figure 4.5). In both cases, individual end-point variance of pointer and targets was larger than their sum, which indicates mutual error compensation, a feature known from motion control synergies (see above). In task space, i.e. the distance between targets and pointer, the movement profiles of all three conditions were almost congruent. The characteristics of cyclical aiming movements, as performed in this study, can also be explained with limit cycle models of oscillatory systems [117,237]. The predicted tendency towards harmonic oscillation for low indices of difficulty could be reproduced congruently for all three conditions of Mottet et al.'s study.

Mottet et al. also observed several differences of behavior and performance between bimanual and interpersonal cooperation. For example, asymmetric role distributions could not be observed in the bimanual condition, but only between two users. In the former, the movements of targets and pointer each contributed approximately half of the error reduction, hence both hands were coordinated in symmetric antiphase.

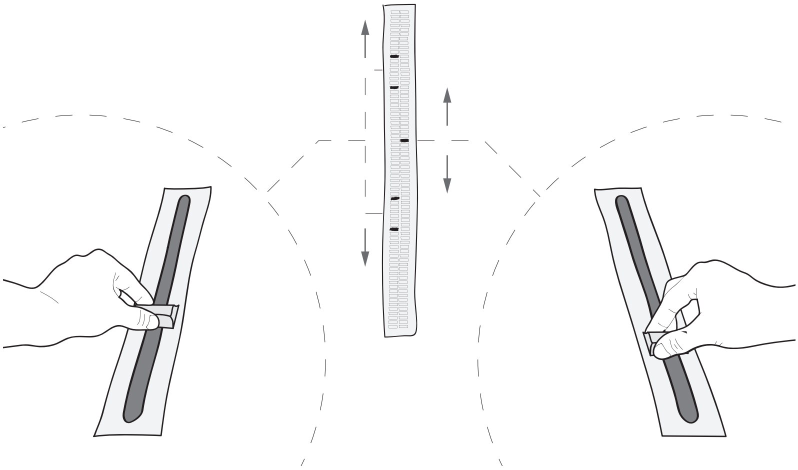


Figure 4.5: *Mottet et al. compared single-handed aimed movements with the coordinated performance of two hands of one or two persons [238]. Their experimental apparatus consisted of two linear sliders and two adjacent vertical arrays of light-emitting diodes (LED). On the left array four active LEDs indicated the extent of two target positions. The position of these targets could be adjusted with the left-hand slider. The LED-array to the right showed the target positions with a single active LED. The target position could be changed with input from the right-hand slider. In four different conditions, the sliders were either operated separately or simultaneously with two hands of one or two users.*

During interpersonal cooperation, instead, the target movement contributed slightly less (43,6%) to the joint error reduction. With increasing task difficulty, however, interpersonal movement coordination converged towards symmetric antiphase behavior. The improved coordination also enabled significant better performance of two cooperating users for tasks with the highest index of difficulty in the experiment – presumably, an effect of synergistic error compensation. Single users, instead performed slightly worse in the bimanual condition, which the authors attributed to a possible coordination overhead of antiphase motion.

Interpersonal movement coordination based on visual perception seems to be robust, as long as visual attention on each other's actions is provided. Groups, often employ additional means of coupling to improve their coordination. For example, dancing pairs build on tactile signals, an external rhythm, and trained choreographies. Orchestras improve synchronization through a combination of mutual awareness, predefined procedure and rigorous conducting. Ad-hoc coordination can often be facilitated through physical linking. Gentry et al. for example, showed with a one-dimensional rotation task that cooperative manipulation of a common haptic input device, can increase target acquisition performance by a factor of $\sqrt{2}$ [107]. Tandem bicycles, two-man crosscut saws, and the rope in a tug-of-war are everyday examples of such tangible connections. Similar devices have also been suggested for increased coordination of collaborating users in immersive virtual reality [6,297].

4.4.1 Joint Action

The research reviewed so far has concerned almost solely spontaneous coordination in simple tasks like finger tapping and target acquisition with few degrees of freedom. The temporal synchronization of involved actors could be explained by the earlier mentioned model of coupled oscillators – a process known as entrainment [170,303]. These couplings may be largely dependent on interlimb synergies that result from attention rhythms, e.g., eye movements while tracking a moving object [194,303]. The question remains, how entrainment affects more complex cooperative action.

Knoblich et al. recently published a comprehensive review of research results on interpersonal coordination and joint action [180]. They distinguished between emergent and planned coordination. Low-level perception-action couplings like entrainment contribute to emergent coordination. Planned coordination, instead, specifies higher-level interaction goals through common task representations and perspective taking. Complex cooperative performances clearly involve coordination on both levels. Knoblich et al. argued that planned coordination builds on the fast perception-action couplings, which in turn are affected by one's goals and interests through conscious attention.

Besides entrainment, Knoblich et al. suggested three further sources of emergent coordination: joint affordance, perception-action matching, and action simulation [180]. Joint affordance recognizes that a group of people has extended capabilities and different action requirements, which affects the affordances of objects, environments, and situations. An open door to a building, for example, invites a single person to enter directly, while it does not afford for a whole group to pass simultaneously (Figure 4.6). Perception-action matching and action simulation are closely related. The former describes the activation of action representations corresponding to perceived ones. The represented actions must not necessarily be performed, but apparently, one is generally better prepared to perform the same actions as those observed in the behavior of others. The behavior is thus also termed action imitation. The internal representation of perceived actions provides the basis for mutual action simulation. Predictions of each other's actions and their effects can be derived that facilitate appropriate reactions, mutual error compensation and the planning of higher level sequences.

The effects of perception-action matching are very similar to those described by stimulus-response compatibility. Studies have demonstrated that people could start a particular movement more immediately if they have just been observing a similar one (e.g. [45,330]). The conceptual difference between both theories is the dimension

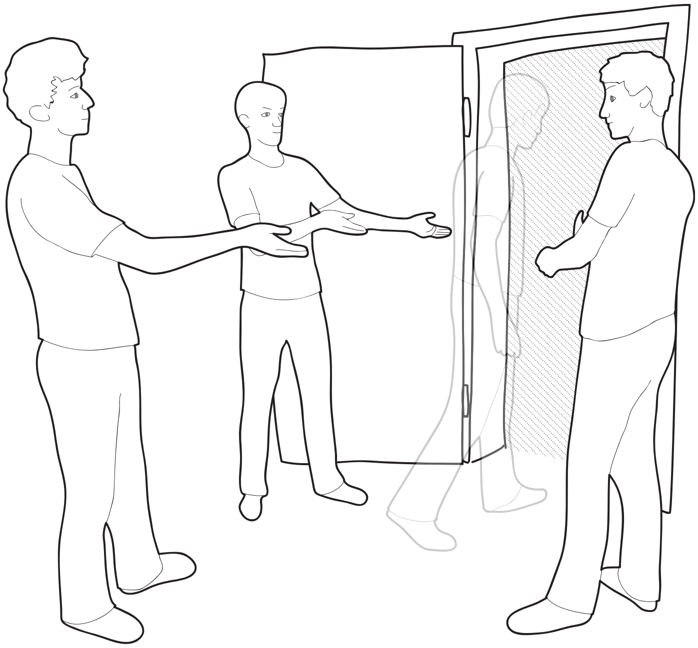


Figure 4.6: *For individuals, an open door affords going through. Its joint affordance is rather to negotiate an order of passing.*

of similarity. Stimulus-response compatibility considers geometrical relations, while perception-action matching relates to ideomotor compatibility, i.e., the similarity of movements. Bach and Tipper showed that watching someone kicking a ball led to faster responses with the foot, while the observation of a typing person prepared for rapid responses with finger motion [17]. Stimulus-response compatibility could not explain this effect, while perception-action matching predicts the general activation of limbs involved in an observed action. According to stimulus-response compatibility it should make no difference, whether we are stimulated by a moving object or a hand following the same trajectory. Brass et al., however, showed that videos of finger movements affected the corresponding responses more strongly than abstract representations [45]. Apparently, both effects can reinforce each other. They also found that, in case of conflict, ideomotor compatibility seems to be the stronger cue.

According to Knoblich et al., planning is required to prepare for joint action and accommodate to changing situations as well as to exchange with collaborators and distribute subtasks. It involves the development of shared task representations, an understanding of each other's capabilities and considerations of potential differences between the collaborators' goals and perspectives. The low-level behavioral patterns described as emergent coordination, on the other hand, enable the accomplishment of joint action plans in real time. These seem to be rooted in perception-action couplings, and therefore depend on one's attention, which follows higher level goals.

4.5 Conclusion

Cooperative action occurs on various levels. This chapter has reviewed psychological research on the synergies of connected body joints, bilateral coordination, and joint action of multiple people. Very similar behavioral patterns were found in these apparently different cases. Considering the generality of the coupled-oscillators model [170], it seems reasonable to expect similar results from observations at further scales, e.g. individual muscles, cells, multiple groups, and crowds. In fact, entrainment effects have also been shown for swarms of insects (e.g. [127]). For the purpose of the present work, we are most interested in the coordination of multiple limbs and persons. The requirements of the involved processes may inform the design of versatile and expressive user interfaces.

Cooperation involves multiple interdependent actors with variable contributions. The more actors that are involved, the more degrees of freedom that must be coordinated for a joint performance. The movement flexibility of the human body is a prime example of this issue. Bernstein observed that the interconnected limbs, combine to a complex system involving more degrees of freedom than necessary to describe the resulting motion trajectories of end effectors. This movement flexibility seems to challenge motor control, but apparently, synergies between the limbs enable even higher accuracy, i.e., the variance of individual limb motion is larger than that of their combined effect. The involved limbs compensate each other's deviations. Such synergistic behavior could also be demonstrated for bimanual and interpersonal cooperation.

Temporal synchronization (entrainment) seems to be among the primary organizing principles of cooperative action. Interacting at conflicting rhythms is indeed very difficult. The robust coupling behavior corresponds to other dynamic systems and can be predicted by models of coupled oscillators. The internal mechanisms of this interaction are not entirely understood. The coupling can potentially occur at the level of motor control in the nervous system, e.g. through synchronized afferent signals. However, entrainment and other effects of dynamic coupling have also been demonstrated between different people that coordinated their actions on the basis of mutual visual perception, hence entrainment also occurs as an effect of perception-action coupling.

The coordination of more complex cooperative actions involves planning and the distribution of subtasks. Consequently, joint action is considered a combination of planned and emergent coordination processes. Conscious action planning sets the goals, negotiates subtask distribution, and guides our attention, while the actual cooperative performance is implicitly controlled through perception-action couplings. Similar hierarchical control models have been suggested for bilateral coordination [210]. Complex actions seem to be further facilitated by an asymmetric division of labor. In the reviewed examples of cooperative action, however, the respective responsibilities are rarely predefined or fully separated. Synergistic error compensation may only occur, if the involved parties can affect the same parameters. The different roles emerge and alternate as required in dynamically changing situations.

The similarities of bilateral and interpersonal coordination indicate that both cases do not need to be considered independently in the design of user interfaces for cooperation. Instead, a workplace that offers a variety of access points for the cooperation of multiple users, may also facilitate bimanual interaction. From the observations discussed above we can derive several requirements for such workplaces. Entrainment, a basic building block of coordination, requires the correct spatiotemporal perception of involved actors. Mutual error compensation, an essential characteristic of synergies, requires that the involved actors may somehow affect the same parameter space. Action imitation and prediction, two further sources of emergent coordination, depend on mutual perception and knowledge on the action capabilities of co-actors. This is also true for joint affordance and joint perception, two sources of emergent and planned coordination, which additionally require coherence of the shared interaction space. Last but not least, subtask distribution or division of labor requires that cooperators can assume different, and potentially independent, roles.

Chapter 5

Related Work on Cooperative User Interfaces

Several established user interfaces and many research prototypes support cooperative input from multiple hands, through different modalities, and also by multiple users. In the following we review a broad range of examples from commodity products to research prototypes. The described interfaces are very diverse and may appear unrelated, but, all of them promise increased expressiveness through the involvement of multiple simultaneous activities.

5.1 Bimanual User Interfaces

Bimanual input is an established pattern of current user interfaces. The combination of manipulation input with the mouse or a pen and symbolic mode switching through keystrokes can be considered an example of asymmetric cooperation[1]. Holding a tablet computer in one hand while operating the touch interface with the other involves an asymmetric division of labor similar to handwriting in a paper notebook. The popular multitouch interface for integral manipulation of position, orientation and scale involves simultaneous symmetric actions of two fingers. The most impressive example of skilled bimanual input is probably touch typing, where 2-10 fingers perform symmetric tapping tasks in phasic alternation.

---

[1] One of the earliest instances is perhaps Sutherland's Sketchpad from 1963 [331]

5.1.1 Touch Typing

One and a half centuries after the invention of the typewriter, the keyboard is still one of the most effective user interfaces. Professionals often achieve typing speeds of 60-90 words per minute with intervals of only 100-200 ms between successive keystrokes [106, 210, 289, 312]. Considering an average word length of 5.1 letters for English[2] and the necessity of whitespaces, this corresponds to 6.1-9.15 keystrokes per second. For comparison with user performance in common single-target acquisition tasks, we can estimate the throughput rates in such consecutive target selections. The button size of a full-size keyboard is approximately 1.5 x 1.5 cm and no letter is further away from the respective finger's home position than 2 cm. Professional typing rates thus correspond to a throughput of 8.6-12.9 bits/s.

This compares roughly to single target acquisition performance as measured in Fitts's original experiments [87]. Seemingly, symmetric input from two hands and multiple fingers does not necessarily increase single target acquisition performance, but it enables very high throughput rates over long durations. The continuous alternation of involved fingers reduces fatigue. Touch typing involves a clear assignment of keys to each hand and finger, which effectively reduces target distances. The specified home positions for each finger on the keyboard enable movement control without visual guidance such that the writer can focus on the text to be written.

Stella Pajunas's world record in 1946 involved one hour typing at a speed of 216 words per minute[3], which would account for an astonishing motor throughput of about 31 bits/s if the actions were serial. These comparisons are even more impressive if we consider that aimed movements are only a small part of the typing performance. Typing involves a choice between multiple keys and an additional cognitive effort for the translation of words to letters and the key positions representing them. How can we realize a similar or even larger number of aimed movements despite this additional effort?

Gentner et al. demonstrated that the sheer keystroke performance with intervals of 124 ms on average is achieved through the parallel motion of multiple fingers, which take more than twice the time individually (261 ms on average) [106]. Flanders and Soechting later confirmed this parallelism [89]. Subsequent movements are initiated before others are completed. Hence, the aimed movement related to each individual keystroke is not faster than others. The improved performance is achieved through the simultaneous cooperation of multiple fingers. The cognitive processes governing the coordination of this cooperative action are not entirely understood, but compelling theories have been developed. According to recent research, main factors include hierarchical task representations, learned associations between stimuli and key locations, and the integration of serial order (of keystrokes) in rhythmic sequences [210, 289].

---

[2] http://www.wolframalpha.com/input/?i=average+english+word+length

[3] see http://www.owled.com/typing.html

5.1.2 Bimanual Spatial Manipulation

Guiard argued that skilled manual actions generally involve both hands in an asymmetric division of labor. The non-dominant hand defines a spatial reference frame for more fine-grained actions of the dominant one [115] (see also Section 4.3). When slicing bread or vegetables, for example, one hand operates the knife, the other one holds the piece to be cut. Also the bimanual operation of tools generally follows this asymmetric division of labor: if the tool is meant to be operated with a forward thrust along its elongation, e.g., a spade, a drilling machine, or a billiards cue, the non-dominant hand stabilizes its posture and orientation at a more distal position, while the dominant hand exerts the necessary force. If the tool is instead meant to be operated with lateral motion, e.g. a golf club, a hammer, or a rake, we prefer to hold it the other way around to enable more controlled and forceful lateral movement through the dominant hand (Figure 5.1).

Figure 5.1: *The bimanual operation of mundane tools is often asymmetric. Forward thrust, e.g., when using a spade, is generally performed with the dominant hand at the far end of the tool handle. The non-dominant hand supports with guidance at a more distal working point. The placement of hands is usually the other way around if lateral tool movements must be controlled, e.g., when using a pick axe. A pushcart is one of the few devices that are operated with both hands in equal roles, to facilitate balance and apply more force.*

Consequently, most bimanual computer interfaces that have been proposed also build on Guiard's model of a kinematic chain. In the realm of computer interfaces, input from the non-dominant hand has been suggested to control the context in which selection and manipulation input of the dominant one is meant to be applied. Examples include the view at a digital document or a virtual environment (e.g. [135, 375]), the handling of a 3D object for closer examination and editing (e.g. [137, 138, 314]), the definition of motion constraints like a pivot point for rotation and scaling (e.g. [62, 67, 173, 314, 375]), or the provision of additional functionalities via menus or magic lenses (e.g. [38, 314]).

Unfortunately, only few of these developments are available in commercial products. The research prototypes used different input sensors or tracking markers for both hands, e.g. touch and pen [44, 143] which are often missing on commodity devices. More recently, devices with touch and pen input have become available, but unfortunately, the combination of both inputs is not yet well supported. They are often interpreted equally and mutually exclusive.

Symmetric bimanual interaction techniques have also been proposed. Most of these were developed to increase the number of simultaneously available degrees of freedom. Established single-pointer input is limited to the selection of coordinates and translation input. With more than one access point manipulated objects can also be rotated, scaled, and even deformed in a single coordinated action (e.g. [67, 100, 135, 200, 201, 203, 208, 236]). Note that also in case of such symmetric interaction of both hands, a tendency towards asymmetric operation can be observed. If it comes to fine tuning the manipulated parameters, the hands alternate between asymmetric roles of holding a reference position and motion input.

The combination of rotation, scaling, and translation input (RST) with two contact points has become a basic building block of multitouch interfaces for 2D and 3D geometric manipulations (e.g. [181, 280]). On mobile touch devices, these gestures are often performed with two fingers of one hand. Whether two hands are employed or two fingers of one hand primarily depends on the size of the screen and the manipulated content. Asymmetric touch input, however, is less established. The following subsection discusses research efforts to facilitate asymmetric bimanual interaction with multitouch devices.

Asymmetric Multitouch Input

Multitouch input sensors implicitly provide a consistent spatial reference frame for motion input from multiple fingers. If the sensors are applied as a display overlay, they even blend motor and display space to a coherent whole. Therefore, these devices are well suited for various bimanual interaction techniques. Surprisingly though, existing multitouch systems make only limited use of these capabilities. Only the pinch-zoom gesture has become an integral part of commercial multitouch interfaces. It is even more surprising that asymmetric interaction is not well supported, as it is more common in real-world settings [115].

One reason for this situation is certainly the context of use. Multitouch sensors are most established in mobile computing, where one hand is commonly required for holding the device. Thereby it already provides a spatial reference frame for input from the other hand. Corresponding to the often-cited example of writing on a sheet of paper, the non-dominant hand holds the medium on which the dominant hand is acting. With digital media, however, the held device is only a physical frame for dynamic digital content. The dominant hand navigates through the displayed information instead of editing it as in the case of handwriting.

For asymmetric bi-manual interaction with the digital content, the non-dominant hand must also provide a reference frame in the virtual interaction space. Several research prototypes have shown that this can be feasible despite the limited motor capabilities while holding a mobile device [95, 133, 224, 352]. Recent developments also showed how device motion and touch input from the same hand can be combined for more efficient navigation input. Integrated inertial sensors complement touch screen input with input options related to the selected on-screen target. For example, Hinckley et al. show how tilting the device can perform a zoom into a map region selected by touch [140].

Asymmetric bimanual touch input requires the discrimination of fingers and hands as a prerequisite for the assignment of different roles. This is also true for stationary devices, where both hands are fully available for cooperative actions. Touch sensors generally cannot discriminate different hands or fingers. Marquardt et al. experimented with finger-worn markers to explore the potential applications if each finger can be identified [220]. Holz and Baudisch realized a touchscreen as a large-scale fingerprint reader [146]. For more widely applicable hardware systems several discrimination heuristics on the basis of fingertip shape and hand posture have been suggested [13, 70, 354, 382].

Alternatively, applications can provide additional input modes for asymmetric bimanual interaction via dedicated buttons and on-screen widgets (e.g. [62]). A more versatile solution is the combination of pen and direct touch input. The explicit discrimination of roles based on the used devices is robust and supports a wide range of powerful cooperative actions [44, 136, 143, 266, 374]. Research on bi-manual interaction performed as part of this thesis showed that asymmetric bi-manual input can also be implicitly derived from the fundamental behavioral pattern that the reference-providing hand generally initiates the cooperative action (see Chapter 9).

5.2 Multimodal Interaction

Interactions with mundane objects, our environment, and other people involve our entire body, not only hands and fingers. We walk towards places, we bend our bodies around obstacles, we look at points of interest, and we specify meaning with gestures and words. Perhaps, the most complex combinations of actions are involved in the exchange with each other. Our verbal expressions only reveal their full meaning in relation to the spatial and temporal context, our posture, our gaze, our facial expressions, our voice and many other behavioral aspects. Computer workplaces, on the other hand, evolved from desk work, which is most often solitary and bound to the interaction capabilities of our hands. The combination of modalities with complementary capabilities promises more flexibility as well as higher interaction efficiency and fluency.

Martin distinguished five types of cooperation between modalities [61]. *Transfer* denotes that information provided through one modality is further processed through another one, e.g., to resolve inaccuracies of individual modalities. *Equivalence* means that a certain process can be realized through various modalities, which provides more flexibility for dynamically changing settings and situations. *Specialization* exploits the different capabilities of various modes. Interaction processes are thus performed with the most suitable available modality to increase efficiency. Cooperation can also be achieved through *redundancy* of multimodal information processing. Simultaneous input through multiple modalities can, for example, provide confirmation without the common and often annoying confirmation dialogue for irreversible actions. The cooperation pattern of multimodal *complementarity* signifies that information processed through different modalities is merged to achieve a more complex expression. It differs from transfer in that the processed information does not only build on each other but it is more tightly intertwined as in the often mentioned combination of manual object selection with vocal commands.

5.2.1 Hand and Voice

The most popular example of multimodal interaction is the combination of pointing and speech input. Bolt demonstrated in 1980 how well this natural combination suits the common interaction pattern of command application to selected objects [41]. Since then, many research prototypes based on this concept have been realized (e.g. [42,254,313,339,341]). With improved robustness of speech recognition and a higher diversification of computer usage scenarios, speech input is also gaining acceptance in commercial systems.

Oviatt emphasized that speech input reveals the largest benefits only in combination with direct manipulation [255,256]. Pointing and tracing, for example is suitable for the specification of spatial parameters. Speech, on the other hand, facilitates the integration of semantic information. Semantic selection filters can resolve ambiguous or erroneous input [256,339]. The verbal expression "Highlight that red car", for example, provides two selection filters and specifies an operation that only requires coarse pointing towards the location of the desired object (Figure 5.2). Schnelle-Walka and Döweling identified successful integrations of speech input to touch-based interaction and developed a taxonomy of related interaction patterns [305]. Their taxonomy includes speech-based mode switching, verbal selection of distal objects, touch-based error correction for speech input, and the above mentioned combination of pointing-based object selection with verbal commands.

Figure 5.2: *Speech and deictic gestures complement each other. The combination of pointing and a verbal description, e.g., "blue book", can clearly disambiguate a target in dense environments like a bookshelf. Moreover, verbal commands can describe meaningful actions with the indicated object, e.g., to read the book.*

Combinations of hand and voice input have not been adopted in end-user interfaces, despite the fact that potential benefits have been known for several decades and voice recognition is ever becoming more robust. Perhaps the reason for thit is that the low complexity of most graphical user interfaces does not require more sophisticated input. For more complex applications in the field of computer aided design, for example, the benefits may be more relevant [42,313]. Another aspect is certainly the social compatibility of multimodal interfaces. The more of our communication modalities become engaged in the dialogue with the machine, the more they will interfere with interpersonal communication. Office colleagues generally talk with each other while working manually on different and often unrelated tasks. Incoming telephone calls interrupt these conversations - after a noticeable jingle that provides mutual awareness. Speech input to the computer would most likely inhibit such social exchange in office spaces.

5.2.2 Gaze and Hand

The combination of manual input with gaze tracking has also received considerable attention in research on human-computer interaction (e.g. [79, 323, 324, 380]. Gaze generally serves for the coarse identification of a target area and thereby provides a spatial frame of reference for subsequent location refinement with manual input. Stellmach and Dachselt expressed this pattern with the catchy phrase "gaze suggests, touch confirms" [323]. Our eyes continuously scan the environment and do not steadily remain on an object or area of interest. Efficient use of gaze input can thus only be realized with such combinations.

Earlier experiments of Zhai et al. [380] as well as Drewes and Schmidt [79] showed performance benefits of this input combination. Zhai et al. reported performance benefits of about 14% for the large-distance cursor movements. Drewes and Schmidt even showed that gaze-supported target acquisition can almost eliminate the effect of target distance. Moreover, they highlighted the impact of visual distraction on target acquisition with relative motion input from the mouse. In a condition with complex background texture they observed target acquisition benefits of almost 33%. Both studies combined absolute area selection based on eye tracking with position refinements from relative pointing devices. This seems to be the only feasible combination, since eye-tracking is well-suited only for absolute input. Relative input from the hand can add accurate adjustments in the final closed-loop phase of target acquisition. When using eye-tracking for relative input, instead, one would immediately lose track of the cursor.

It should also be noted, that in neither of these studies on long-distance target acquisition were users able to achieve throughput rates beyond approximately 3 bits/s[4]. In comparison to common 2D pointing performance, this not very convincing (cf. Section 2.1). Zhai et al. discussed this issue and suggested that the isometric joystick employed in their studies could be the reason for the overall mediocre performance [380]. Another possible explanation is that such gaze-supported manual pointing requires specific training, since its behavior contradicts established perception-action couplings. We are used to follow the actions of our hands with our eyes or to move our hands to locations we are looking at, but, we do not expect that manually operated tools automatically follow our gaze, while we are not even moving the hands.

Research results on the combination of gaze with manual input are far from being conclusive. Potential interferences of gaze input with cognitive tasks and social interaction have not yet been explored. Also, the potential performance benefits are

---

[4] The data from Drewes and Schmidt allows the computation of an average throughput of about 2.3 bits/s for both input conditions in case of a blank background [79]. In Zhai et al.'s study, the manual input condition using an isometric joystick resulted in an index of performance of 3.2 bits/s [380] or a throughput of about 2.7 bits/s (average index of difficulty divided by average movement time). In combination with gaze input, a maximum throughput of 3.13 bits/s could be achieved.

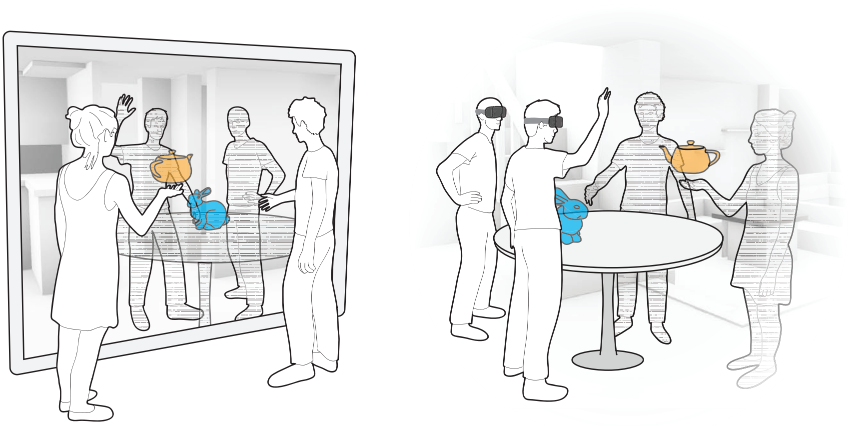
not yet fully proven. Improved target acquisition has been shown (e.g. [79,323,380]), but, similar or better performance can be achieved with other input combinations or clever input mappings (e.g. [93,97,226,244,350]). Pfeuffer et al. recently suggested a combination of gaze input with multitouch and pen, which produced comparable performance to the combination of pen and touch [266]. Nevertheless, in a study by Nancel et al., users expressed their subjective preference for combinations of gaze and hand and achieved higher accuracy than with combinations of manual input modes only [244].

5.3 Multi-User Cooperation

The direct cooperation of multiple computer users has been explored in collocated and remote settings. Collocated collaboration systems are often built around tabletop displays (e.g. [179,233,295,310,333,336] or large public screens (e.g. [16,50,156,160, 262]), and they often offer combinations of multiple displays (e.g. [243,332,353,365]). Most of these collaborative systems use 2D displays, but also multi-user stereoscopic displays have been suggested (e.g. [4,99,174], Figure 5.3 a). The latter can create the visual perception of virtual 3D objects in a shared interaction space. This requires to produce individual images for each eye of all involved users and the number of simultaneously perceivable images on shared displays is generally limited. Therefore, head-mounted displays are more commonly used to create a shared virtual or augmented reality for collocated collaboration (e.g. [47,177,297,300], Figure 5.3 b).

Also systems for remote collaboration have been studied extensively. Most commonly, abstract representations of the participants and their activities in virtual 2D and 3D interaction spaces have been suggested (e.g. [53,101,122,144,285,309,343,363]. For interpersonal communication, however, realistic representations of collaborators can be beneficial. The possibilities and challenges of 2D and 3D videoconferencing has received considerable attention in that regard, in particular, the technical realization of eye contact [18,23,129,155,162,232,239,245,308,349]. More recently, collaborative virtual environments that build on real-time 3D capturing of people to provide highly realistic avatar representations has become a vibrant field of research [26,30,112,218,265,346,383] (Figure 5.3).

The primary goals of most collaborative user interfaces are: a joint experience, mutual awareness and the efficient exchange of information. People meet to discuss their ideas and viewpoints on a certain topic and often aim for a common understanding. Harrison and Dourish suggested to distinguish the semantic qualities of a shared place from the geometrical attributes of a shared space. "A place is generally a space with something added – social meaning, convention, cultural understandings about role, function and nature and so on." [128]. They argued that the behavioral properties of a meeting place are more relevant than its spatial configuration and that the emergence of a shared sense of place primarily requires "support for adaptation and



(a) Multi-user 3D screen (b) Head-mounted displays

Figure 5.3: *3D displays can create the visual perception of virtual 3D environments. Multi-user 3D screens (a) and head-mounted displays with see-through capabilities (b) allow to share this experience with others. The image on the left shows two users in front of a large multi-user 3D screen. They both see a virtual teapot (yellow) floating above the hand of one user. The image on the right shows two users with head-mounted displays that can see a virtual bunny model (blue) augmented on top of a physical table. Telepresence technologies support real-time 3D capturing and reconstruction, such that people from different locations (here: image a and b) can meet in a shared virtual environment. The appearance of their virtual avatar representations is illustrated with a semi-transparent hatched texture.*

appropriation”. Consequently, places for mutual exchange can also exist without any spatial attributes, e.g., social websites. Nevertheless, the spatiotemporal coherency of the shared space may contribute to the efficiency of communication. A shared interaction space, as available in collocated settings for example, facilitates interpersonal coupling. It allows users to coordinate their actions and illustrate their views through body language and the exchange of reference objects. The improved communication fluency can also support the emergence of a shared sense of place.

Buxton distinguished between the task space and the person space in collaborative work and argued that “effective telepresence depends on quality sharing of both” [49]. In collocated single-display settings, this is generally a given, since all users share a coherent interaction space, whereas in telecommunication setups, these two spaces are often dissociated. In video telephony, for example, we may share an application screen for taskwork while we view our collaborators in additional windows. Buxton argued, that both spaces do not need to coincide. The hydra system, for example, represented each participant as a miniature video terminal that was spatially unrelated to a shared screen [49]. The system, nevertheless, supported a shared person space where the focus of attention was apparent in the participants' orientation towards individual video terminals that represented the other participants or towards the task shown on the shared screen. If the person space and the task space do not coincide, however, it is difficult to maintain awareness of action authorship in the shared task space. Ishii and Kobayashi's Clearboard [155] showed a more coherent combination of the person and task space in remote collaboration. Following the metaphor of a transparent white board, their system created a coherent workspace between two distant collaborators.

Gutwin and Greenberg considered workspace awareness, defined as the "up-to-the moment understanding of another person's interaction with the shared workspace", as a major requirement of collaborative systems. They emphasized three main elements: 1. the presence, identity, and authorship of participants (who), 2. the involved artifacts, actions and intentions (what) and 3. the location, gaze, view and reach of users (where) [121]. Moreover, the framework includes the history of artifacts and events (how and when). They noted that the ease of people maintaining workspace awareness in real-world collaborative settings is based on the continuous gathering of this information through consequential communication (the observation of each other's activity), feedthrough (the sensory perception of involved artifacts) and intentional communication (verbal and gesturing). Despite the advantages found for the rich information exchange in the real world they showed that workspace awareness can also be achieved with more abstract means in groupware for distributed collaboration.

5.3.1 Manipulation Conflicts

As humans, we do not only meet to talk and gesture, but also to explore the world together and adapt it to our needs. This is no different in virtual environments fro collaboration. Joint object manipulation has received particular interest in research on collaborative user interfaces – often due to observed coordination issues. If several participants are provided with equal means to manipulate objects in a shared environment, their actions may cause mutual interference rather than support (e.g. [235, 262, 282]). Such conflicts seem to be particularly pronounced if the input and output spaces (person space and task space) are not overlapping and participants may intrude each other's personal territories [149, 156, 268] (Figure 5.4).

In real-world settings, similar situations occur – and in case of many games this struggle for control over a particular item is often the whole point of the action (consider ball games as an example). However, we can generally build on experiences with similar objects, in similar setting, and under the same physical laws, which provide us with an implicit understanding of each other's action capabilities. We can anticipate the goals and actions of others, since objects must be in reach for manipulation and we can observe movements towards the areas of interest (see Section 4.4).

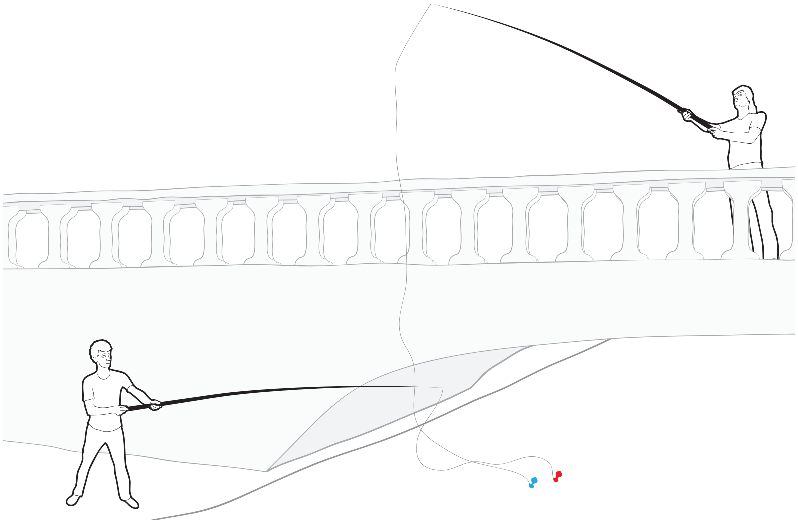


Figure 5.4: *Research on multi-user collaboration in virtual environments often emphasized the issue of manipulation conflicts and suggested techniques to support the negotiation of access. In real-world settings people successfully coordinate their actions based on mutual observation and synchronization. The spatial dissociation of user input and its effects, as it often occurs in computer applications, hinders the necessary awareness. If the person space and the task space cannot be perceived as a coherent whole, people may unconsciously violate each other's interaction territories.*

Some of these cues can also be exploited in collaborative computer applications, if the person space and the task space are consistent, i.e., if the spatial relations of user actions and their effects are directly related and observable. With multitouch displays, for example, input can be directly applied to the visible application content, while movements of a mouse pointer are spatially separated from the operated input device. Hornecker et al. compared these two paradigms for multi-user interaction on a tabletop display [149]. They observed more concurrent interactions in the touch condition which allowed for mutual support but also increased the frequency of conflicting actions. With multiple mice, instead, users tended to act one after another, in order to permit the monitoring of each other's actions and avoid potential interferences. The higher mutual awareness in the touch condition facilitated negotiation and conflict solving, which allowed them to take more risk and engage simultaneously in the cooperative task.

Collaboration often requires to anticipate the actions of others. In real-world settings, we can build on our experience of possible actions under invariant physical constraints. This expectation, however, does not hold in the context of computer interfaces. Objects can be manipulated from any distance and any type of user action can be mapped to the various effects. The manipulation of virtual objects is not constrained by physical laws and the consequences of concurrent user actions need to be explicitly considered. Furthermore, distributed systems for remote collaboration suffer from synchronization issues: Users at different locations may want to affect the same attributes of an object simultaneously without being aware of their concurrent actions. Greenberg and Marwood identified this concurrency control problem in distributed groupware and showed that neither locking, nor merging, or forced serialization of user input can solve all of the resulting problems [111]. They discussed the tradeoffs related to various implementations of these concurrency control schemes and proposed to apply user-centered design methods to find the best compromise for each specific use case.

In collocated groupware, typically system consistency is maintained by managing the application state through a single application and the physical presence of involved users improves their mutual awareness. However, Morris et al. found that social protocols are not always sufficient to prevent conflicts [235]. Unexpected side effects of interaction methods are a part of the problem. Ease of use can become another issue. Consider common actions like maximizing a GUI element or panning the workspace. Direct access to global transformations is clearly beneficial for single user workplaces, but in the context of collaborative applications, one must keep in mind that the results affect others too. Gutwin and Greenberg found that the usability requirements for individual users and groups can be conflicting [120]. In the context of remote groupware developments, they suggested to enable individual views of the shared interaction space and explicit indicators for mutual awareness to alleviate these issues.

Morris et al. also observed conflicts in collocated settings and discussed the design space for application-controlled coordination policies both for global changes and individual object transformations [235]. They suggested a distinction of policies based on the source of the coordination initiative. Proactive coordination relies completely on the decisions of the user who initiated the conflicting action. Reactive policies consider only the situation of everybody else and mixed-initiative coordination techniques combine aspects of both policies. Their systematic analysis of the topic provides a good starting point for specific groupware developments. Following Greenberg and Marwood's considerations on distributed groupware [111], they argued that there is no ideal solution, but the suitability of various approaches depends on the users and application scenarios.

5.3.2 Cooperative Object Manipulation

Researchers have also explored the combination of input from multiple users in virtual reality scenarios – often following the example of joint object handling in the real world. Similarly, users can pick large virtual objects at multiple points to simplify their handling while maintaining the accuracy of the larger scale (e.g. [90, 101].

In the real world, however, the object provides a physical link between all involved hands which helps to coordinate their individual motion through haptic guidance. The physical link also constrains the motions of each hand. During the manipulation of virtual objects such a physical link is missing. Interaction techniques need to map the unconstrained movement of the involved hands to the available degrees of freedom of the object. There is no simple solution to this problem that is intuitive to the users and also results in a controlled object motion. A pseudo-physical or physical simulation of the object manipulation can improve the sense of control by providing feedback for the deviation of the hands from their initial grip positions with rubber-band visualizations [5, 100] or bending input widgets [282]. However, Salzmann et al. [297] and Aguerreche et al. [6] found that a shared tangible input device for multiple users can be preferable.

Hornecker and Buur suggested further benefits of tangible interaction for collaboration [148]. Besides the already mentioned benefits of physically perceived coupling through *tangible manipulation*, their conceptual framework includes the themes of *spatial interaction*, *embodied facilitation* and *expressive representations*. *Spatial manipulation* emphasizes the relevance of a shared interaction space in which the movement of objects and one's body has a comprehensible meaning. Support for full body interaction encourages performative action and may thereby leverage body language. User interfaces should ensure that everybody can continuously follow the interaction process and avoid *fragmented visibility* as much as possible. *Embodied facilitation* highlights that tangible interaction devices embody *physical constraints* and provide *multiple access points*, both of which can balance the involvement of participants. The physical characteristics of environments and objects afford particular usage patterns. Moreover, *tailored representations* of interaction devices can better fit the users' skills and experiences. The latter is closely related to the concept of *expressive representation* which involves meaningful and long lasting representations of functionality and content as well as comprehensible interrelations between physical and digital artifacts (*representational significance* and *perceived coupling*). The authors further argue that tangible objects can facilitate the immediate externalization of ideas and provide unambiguous references for the communication among group members. Three case studies confirmed the benefits of the specified features of tangible interaction for multi-user cooperation.

The cooperative manipulation of objects generally involves two or more contact points, which implies the simultaneous manipulation of position, orientation and, if permitted, also scale. Such highly integrated control may facilitate coarse approximation, but it complicates accurate adjustments of individual degrees of freedom. The required coordination for co-operating users acting on the same degrees of freedom of a single object is challenging. Clever input integration schemes could alleviate this issue. Ruddle et al. explored the effect of symmetric and asymmetric input integration on the performance in a collaborative 3D manipulation task [293]. In the symmetric case, only motion input that is induced by both users will be applied. Asymmetric input integration, instead, applies the mean of both actions. They observed that symmetric input integration is preferable if both users aim for the same motion trajectory, but that the asymmetric conditions led to better results during periods when the task required both participants to move in different directions. In both conditions, however, they observed a significant cooperation overhead.

In the context of a similar 3D manipulation task, Pinho et al. suggested to distribute the control of multiple degrees of freedom (DOF) among participants [269]. More specifically, they assigned the control of position and orientation to different users and also experimented with a separation of motion in depth from motion along the image plane. This approach of distributing DOF among users had earlier been proposed for cooperative object manipulation in 2D user interfaces [46]. In both cases, the distribution of DOF enforced collaboration. Pinho et al. reported that the separation of input parameters can increase manipulation accuracy, and that the cooperative operation allows users to control more degrees of freedom simultaneously. They also observed, though, that the distribution of DOF negatively affected the comprehensibility of the interface.

Benford et al. proposed to encourage collaboration with a variety of simultaneously available tools that can be used in combination to realize additional functionalities [27]. Instead of enforcing collaboration, this approach is promising increased efficiency, fluency or fun by means of collaboration. They developed collaborative storytelling applications for children that were offering a variety of digital tools for simultaneous operation by multiple mice. With initial implementations, a tendency towards individual action was observed among their test users as well as competition instead of collaboration. As a result of this observation, the researchers implemented extended functionalities that could only be achieved collaboratively, e.g. mixing colors from two differently colored drawing tools, to encourage more cooperation. Children that were testing the revised applications appreciated the collaborative tool behavior. It still seemed difficult, though, to discover the combined features and use them for a common goal. Benford et al. concluded that opportunities for cooperation should be clearly visible and indicate concrete benefits.

Morris et al. later explored cooperative touch gestures using a multitouch tabletop display with user identification [234]. Their collaborative drawing application proposed cooperative actions to realize 1. implicit agreement on global state changes, 2. simplification of effortful actions like reaching over the table, and 3. playful manipulation of larger parameter sets through simultaneous actions. Test users endorsed cooperative gestures to express their agreement to global changes but expressed little understanding for required cooperation if the same result could be achieved alone with only slightly more effort. As an example for the simultaneous manipulation of parameter sets, the test application included an input gesture to manipulate the intensity and thickness of a stroke while it is drawn by another user. The users disliked this distribution of control over the appearance of the stroke. They complained about confusion, inefficiency, and tedium due to the artificial separation. Moreover, Morris et al. reported that several test users felt uneasy about cooperative gestures that required intimate proxemic distances in order to express agreement.

Collaborating people frequently switch between tightly and loosely coupled cooperation [20, 161]. These dynamics can be reflected with a flexible workspace structure that fluently adapts to varying needs [154, 333]. Scott et al. observed the pattern of territoriality in collocated collaboration. Territoriality is a behavioral patterns of humans and animals to establish affiliations with particular areas or spaces through regular usage or occupation. In the tabletop setting Scott et al. observed the emergence of distinct interaction areas for personal interaction and group exchange as well as the establishment of further subspace for storage [311] (Figure 5.5). Supporting this user behavior can reduce input interferences and increase the effectiveness of collaborative work [156, 189, 336, 340]. Multiple separate display and interaction devices like smartphones and tablets inherently serve as personal and storage territories, but effective means for exchange with a shared group territory are required to support collaboration (e.g. [227, 243, 365]).

Figure 5.5: *Emergent territoriality in a tabletop collaboration setting. Participants in collaborative interaction implicitly establish separate areas for private activities (green) and group exchange (blue). Also the emergence of storage areas (yellow) can be observed. The latter are often subspaces of private or group territories.*

5.4 Conclusion

This chapter reviewed research and interface developments in the fields of bimanual interaction, multimodality, and multi-user collaboration. The review highlighted the potential of cooperation as a general principle towards more expressive human-computer interaction. Combining activities of two hands, multiple modalities, or several users promises considerable advantages.

Among existing interaction systems and prototypes, bimanual interfaces revealed the most obvious performance benefits. Unfortunately, only few available sensing technologies can distinguish input from different hands. Technologies to distinguishing dominant from non-dominant input roles in asymmetric bimanual division of labor seem to be required for further progress in that direction.

The benefits of multimodal interaction appear to be obvious at first sight. So far, however, prototypical systems do not live up to the expectable leap in performance. Cognitive overheads of suggested multimodal input combinations may constitute a bottleneck. Moreover, we argued that the design of multimodal interfaces should put a stronger emphasis on their applicability in social contexts.

With respect to multi-user cooperation, the literature review focused on real-time interaction in collocated settings and telepresence applications, where the coordination overhead often seems to undermine many potential advantages of cooperative interface prototypes. The success of social web applications, on the other hand, clearly shows many advantages of cooperation with others. The latter enable social information exchange based primarily on asynchronous communication with remote correspondents – partially at the expense of awareness for local situations and collocated peers. It can be argued that the next step in interfaces for social cooperation requires the seamless integration of local and remote peers in joint action.

The motivations, the application setups, and the involved processes differ between the three reviewed fields of research. However, several interaction patterns and requirements seem to be very similar. The following chapter discusses these similarities and suggests common cooperation patterns to derive essential cooperation requirements and high-level design principles for cooperative user interfaces.

Chapter 6

Cooperation Patterns and Requirements

In the last chapter we reviewed a variety of user interfaces that promise expressive user input through the integration of multiple simultaneous input streams. From the comparison of these different approaches we can derive common situations and requirements. As a basic structure for this analysis, we adopt Martin's typology of multimodal interaction patterns [61]. In the following sections, we discuss its extended applicability to cooperative user interfaces in general and the resulting interface requirements in terms of workspace characteristics and the cooperative coupling of participants. Moreover, we suggest high-level design principles that promote expressive user input across various platforms and settings, including the implications of social settings and multi-user interaction.

6.1 Cooperation Patterns

Martin proposed a typology of cooperation patterns for multimodal interfaces [61]. The types of cooperation he identified, namely *specialization*, *equivalence*, *redundancy*, *complementarity*, and *transfer*, can also be observed in bimanual interaction and the collaboration of multiple users. We therefore adopt this classification and extend it where necessary. From such a broadened perspective we can identify relations between these concepts and other cooperation patterns.

6.1.1 Specialization

The impressive example of touch typing demonstrated that large performance benefits can be expected from *division of labor*: the parallel execution of subtasks that would otherwise need to be performed subsequently. Apparently, the cooperative performance can benefit from strict role assignments in order to avoid confusion. In Martin's typology this is clearly a case of *specialization*. Each finger becomes assigned to a subsection of the keyboard. Research on motor control (see Chapter 4) indicates that this type of cooperative action with clearly separated contributions builds on tight *temporal synchronization* and *hierarchical control*. Both seem to be hard-wired coordination mechanisms of our central nervous system, but, the large performance differences between different typists also show that effective coordination requires extensive training.

Similar cooperation behavior based on *specialization* can be observed in different types of collaborating groups. The members of orchestras, emergency crews, and sports teams assume specific roles that reflect their particular skills and also their situational opportunities during a joint action. Similar to the fingers of a typing hand, their coordination exploits mutual synchronization (or entrainment) and hierarchical control (see Section 4.4). They act together to achieve common goals that go beyond the capabilities of individual members.

Collaborating groups often need extensive training together to achieve a smooth and effective joint performance. It is therefore not surprising that several studies on multi-user cooperation have revealed additional coordination efforts and requirements rather than performance benefits (e.g. [235,269,293]). User studies with novel interface technology rarely allow for sufficient training time. Participants need to perform together with unknown collaborators and in unfamiliar environments. The necessary preparation does not only involve learning the particular task, but more essentially, to learn about the skills and behavior of the others as well as to understand each other's roles in the collaborative process. Individual actions need to incorporate the capabilities and the expected reactions of others. The establishment of such coordinated group behavior can take days and weeks of joint action. Heuristic evaluations may thus be the better choice in many cases (e.g. [20,148]).

Cooperation partners, with or without specialized roles, need to coordinate their actions and establish a shared goal. Hierarchical control can facilitate both, but in case of complex interrelations mutual awareness is indispensable for effective coordination. A conductor, a commander, or a trainer can convey the high-level structure of joint actions, but the fluency of a coordinated process also builds on the relative timing and the logical coherence of the intermediate steps. *Specialization* can facilitate the necessary coordination as it is easier to anticipate the actions of others if they assume predefined roles. Mutual awareness can then build on a restricted set of modalities and cues. For example, the musicians in an orchestra synchronize acoustically with each other, but visually with the conductor. Similarly, during skilled typewriting, the higher-level control structures do not need to monitor the actual buttons pressed by each finger, but proprioceptive feedback on finger pose and tactile information on successful execution may suffice.

Specialization highlights the benefits of diversification for cooperative action. It promotes improved performance in individual subtasks. The diversity of complementary contributions can increase the expressiveness of cooperative actions and also task efficiency through division of labor. The assignment of different roles can also minimize conflicts. Predefined roles among cooperating partners support the anticipation of behavior and thereby facilitate mutual awareness.

6.1.2 Equivalence

Successful cooperation does not always require predefined role assignments as implied by the pattern of *specialization*. People with essentially equivalent roles cooperate spontaneously, e.g., to help each other in carrying heavy luggage up a flight of stairs. Often there are many options to support each other, but the cooperation partners need to settle on a single joint plan. Behavioral norms guide the decision, but they often permit many solutions. Nevertheless, it generally only takes seconds to agree on a plan and execute it, without the need for many words. Research on joint action indicates that such emergent coordination builds on the subconscious behavioral mechanisms of joint affordance, entrainment, perception-action matching, and action simulation [180]. Conscious action planning, on the other hand, contributes to an understanding of the task at hand, potential obstacles, and other involved perspectives (see Chapter 4.4).

In terms of Martin's typology spontaneous cooperation generally follows the pattern of *equivalence*. The allocation of subtasks is flexible, which supports dynamic adaptation to changing situations and the respective capabilities of participants, i.e., temporary specialization in an ad-hoc fashion. The spatiotemporally coherent perception of each other's actions and group-related affordances seems to be even more relevant in this case than they are for collaborating groups with highly *specialized* roles. In the latter case, short moments of dissociated or fragmented perception might be bridged if everybody focuses on the proper performance of their own trained action sequences. If the whole cooperative action is unfolding spontaneously, however, a short moment of distraction can easily lead to conflicting task representations (e.g. sidestepping left or right to avoid forthcoming obstacles).

Equivalence requires more improvisation than *specialization*, but in turn, it offers more expressive freedom during the performance (e.g. frequent role switching). A less rigid structure can also better adapt to dynamically changing situations. Moreover, the cooperation pattern of *equivalence* is more inclusive. Hornecker and Buur emphasized that *multiple access points* can motivate people to get involved in collaborative interaction [148]. It allows a diversity of participants to contribute according to their skills and interests instead of predefined roles. In this sense, interface design based on the pattern of *equivalence* supports emergent specialization. However, with an increasing number of participants, coordination becomes more difficult. Therefore, larger groups often build on more predefined *specialization*

In real-world cooperation it can be useful to combine actions of the same type, e.g., to move heavy objects with more strength. The simultaneous contributions are physically mediated through the manipulated objects. In computer applications this is not the case and the concurrency of very similar actions can instead induce conflicts as discussed in Section 5.3.1. Most interfaces that can sense multiple equivalent input actions in parallel (e.g. multitouch) thus only respond to one of the input actions (generally the first one). It seems ineffective to sense multiple inputs but apply only one.

Extended functionalities through combinations of equivalent input, however, require the identification of inputs, e.g. through markers or specialized tools like pen and touch (see Section 5.1.2). It remains a challenge to build interfaces that offer *equivalence* in terms of *multiple access points* but also support a meaningful emergent *specialization* of these parallel inputs based on the constraints of the situation.

Equivalent interaction roles can also be useful to increase success rates in cooperative tasks that involve a certain amount of chance. Additional simultaneous attempts certainly increase the odds to hit a moving target or to find a hidden feature. Cooperation of this type, however, corresponds more closely to Martin's pattern of *redundancy* which we discuss in the following section.

Equivalence refers to equal opportunities of cooperating parties. In contrast to *specialization*, it promotes concurrent accessibility and situational flexibility. This may lead to conflicts, but it also fosters broad engagement and the conscious negotiation of interaction plans (see *redundancy*).

6.1.3 Confirmation and Intensification (Redundancy)

Martin suggested the cooperation pattern of *redundancy* to describe the concurrency of actions and effects with equivalent meaning [61]. Martin refers to exemplary combinations of typing and talking. If both convey the same expressions simultaneously, they reinforce their meaning and imply confirmation. Using vision and audio redundantly for in- and output, can furthermore improve the comprehensibility and learnability of interfaces [61]. Such redundancy serves to intensify an effect or to express confirmation and agreement. *Confirmation and intensification* seems to be a better suited name for this cooperation pattern as it reflects its usefulness.

Confirmation is obviously helpful to minimize uncertainties, but also unequivocal perceptions and indistinguishable expressions can be emphasized through affirmative actions and effects. During everyday interaction with physical objects, we perceive their presence and behavior consistently with different senses that mutually confirm the perceived phenomena. When clinking glasses, for example, we can see, feel, and hear them touching. This is fortunate, because if only visual feedback were present, we would not be able to look into each other's eyes and many more glasses would get smashed at parties.

Interactive computer applications like virtual reality can simulate some of these phenomena, but especially the simulation of haptic feedback is very challenging. The perceptual presence of virtual objects and environments is thus inherently limited and system flaws like latency and noise in the interaction loop have additional detrimental effects [229]. In collaborative settings, multiple users can mutually confirm their experiences. Virtual objects and geometrical features that are not only visible, but also indicated by others and discussed together become more meaningful. The confirmation of visual phenomena in a shared virtual environment can increase their plausibility (Figure 6.1).

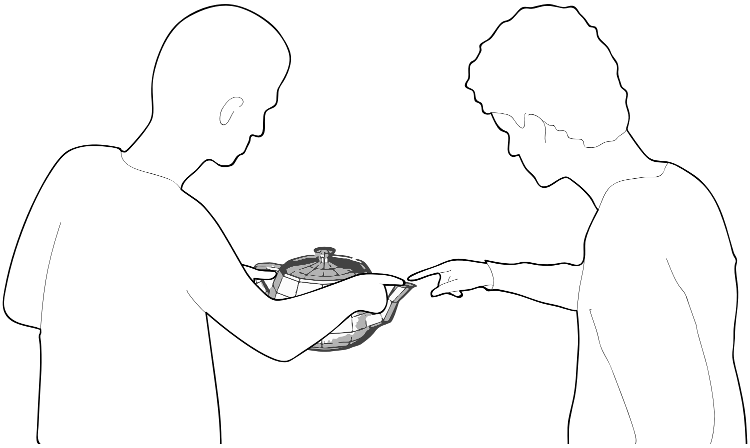


Figure 6.1: *Two users point to the spout of a virtual tea pot. One user's pointing would be sufficient to indicate this geometric feature, but if both are pointing to the same location, they confirm each other's visual perception of the shared virtual scene.*

Multiple users performing the same actions redundantly, can also confirm their agreement in a collaborative process. Morris et al. suggested to extend the scope of input gestures and their effects through simultaneous execution by multiple users [234]. For example, they implemented a global interpretation of an erase gesture in a drawing application that deleted a single stroke if performed alone, but it cleared the whole screen if performed together.

As noted above, performing the same actions simultaneously is also useful in case of limited probability of immediate success. One of our studies on collaborative virtual reality, revealed a significantly increased success rate through multi-user cooperation in a spontaneous visual search task (see Chaper 10). Groups of six users were moved simultaneously through a virtual environment and searched for the same set of objects. They did not had the opportunity to coordinate their actions beforehand, e.g., to focus each on a different subspace, hence, they participated with equivalent roles and tasks. Only through pointing at found items during the ride they could exchange about their findings and increase their individual object identification rates.

The pattern of confirmation through redundant actions has also been suggested in the context of bimanual user interfaces. Wyss et al. presented a bimanual 3D manipulation technique with two pointing devices [370]. Commonly, a single virtual ray is used to indicate an object and its selection is confirmed by clicking a button or switch. The bimanual technique of Wyss et al. omitted the button click event. Instead, it relied on the redundant selection of the same object with two pointers.

Confirmation and intensification builds on *equivalence*. Cooperation partners do not assume different roles but they rather perform the same action in order to express their agreement, or to increase success rates in probabilistic tasks.

6.1.4 Complementarity

Complementarity can be considered the most general cooperation pattern from Martin's typology. It seems obvious that cooperation partners can best support each other with complementary tools and skills. According to Martin's description of *complementarity*, the effects and results of specialized tools and actions are merged to achieve more complex or more specific expressions with less effort [61]. A ruler serves to measure distances, a pen supports drawing. Together they facilitate drawing straight lines.

Complementary actions are generally asymmetric contributions to a joint effort that compensate mutual shortcomings or realize extended effects. During a car rally, for example, the driver is busy to keep the fast-paced vehicle under control. A co-pilot complements the joint navigation with a focus on wayfinding. A baritone, an alto, or a soprano sing at different pitch individually, but their voices complement each other in choral. *Complementarity* builds on *specialization* and it includes the cooperation pattern of *transfer*: the results of an action provide a basis for complementation by others.

Bolt's seminal work on combined input from hand and voice is an often cited example of multimodal *complementarity* [41]. Hand gestures support the disambiguation of objects that are not fully specified by voice, but only indicated by demonstrative pronouns (this, that, ...). Speech can add non-spatial differentiators like colors or object classes, which helps in case of spatial object density and also to specify the requested operation [256, 339]. The command to "align that red cube with the blue one over here" does not make any sense without accompanying spatial gestures in an environment of colored cubes. The gesturing alone would neither be very informative. Their combination, however, enables complex expressions with ease.

Benford et al. explored benefits of *complementarity* in a collaborative drawing application for children. They observed that the *complementarity* of simultaneously available tools can improve the cooperative behavior of collaborating users. Tool combinations that realize extended effects encourage cooperation [27]. Specifically, their experimental application supported the mixing of virtual pen colors to realize a larger color palette. In comparison to the highly effective combination of complementary input modalities, described above, this example may appear less convincing. However, it clearly demonstrates the potential benefits of *complementarity* for the collaborative behavior of multiple users – even with only slightly specialized tools (different colors). Real world examples of musical ensembles demonstrate the expressive potential of cooperation by *complementarity*. No single musician would be able to create the complex sound of an orchestra. The proficiency of collaborators in certain fields or disciplines (*specialization*) apparently leverages *complementarity*.

Complementarity builds on the combination of *specialized* skills and tools. Asymmetric contributions of cooperation partners extend the expressiveness of joint actions.

6.1.5 Transfer

Cooperation is often a sequential process, i.e., something is prepared and further refined in subsequent steps. Martin used the term *transfer* to describe such cooperative processes and referred to the combination of keyboard and voice input as an exemplary multimodal interface. Dictating a text is often faster, but also more error prone than typing, hence the *transfer* from speech to written words and letters for subsequent graphical editing can be advantageous (see [61]). Also, the previously mentioned examples of cooperative input from gaze and hand seem to be most effective if applied according to the *transfer* pattern: gaze provides a reference frame for more fine-grained actions of the hand. In this sense *transfer* can be considered a subset of *complementarity*. Fast but imprecise actions, for example, become complemented with slower but more accurate input.

Guiard's kinematic chain model of bimanual cooperation also follows this general pattern: the dominant hand operates with higher frequency and accuracy in a reference frame provided by the non-dominant one [115]. The *transfer* in this case occurs between different reference frames of motor control instead of different modalities. The combination of locomotion or virtual view navigation with manipulation also corresponds to such a concatenation of reference frames. Any manipulation input is applied relative to one's current position. If navigation and object manipulation is distributed between different participants in a collaborative setting, they also cooperate according to this pattern of *transfer*.

In Martin's typology, *transfer* was meant quite literally as an information transfer between different interaction modalities such as voice, gaze, and motor actions. The described cooperation, however, can be conceived more generally as a *concatenation* of complementary actions and events that build on the intermediate results of one another. Evidently, the mutually provided substructures or references are only meaningful if the cooperating parties share a coherent reference system, which can include spatial and/or semantic dimensions. In the above mentioned case of multimodal text production, the shared reference space is language as perceived through written letters and words. The kinematic chain in bimanual cooperation, instead, describes spatial relations that are perceived through proprioception and vison. Multiple collaborating users can also provide each other with spatial and semantic references, but they do not share each other's proprioception. The exchange of spatial references among different people thus requires external sensory feedback, e.g., mechanical coupling through a haptic link or visual information in a coherently perceived, shared interaction space.

Transfer is an interaction pattern that describes successive processing and refinement of an expression at subsequent stages. It also describes the concatenation of actions and effects as it occurs in a kinematic chain.

6.2 Workspace Requirements for Cooperation

We have seen that spatiotemporal coherence is a prerequisite for fluent cooperative action. Workspace awareness, i.e., the continuous understanding of the presence and activity of others in a shared environment, emerges from many subtle cues in real-world settings [121]. Inconsistencies in the spatiotemporal relations between cooperating partners, their activities and resulting effects make such an understanding much more difficult. Moreover, the building blocks of emergent coordination, i.e., entrainment, joint affordance, perception-action matching, and action simulation, heavily rely on perceptual coherence (see review of [180] in Chapter 4.4). Cooperation by *redundancy* and *transfer* seems to be particularly dependent on a joint reference system, but also the ad-hoc coordination in cooperation by *equivalence* requires continuous awareness and comprehension of others and their actions.

User interfaces for remote interaction and collaboration imply spatial inconsistencies because the effects of input actions are applied elsewhere. Moreover, the processing and transmission of information takes time, hence, most user interfaces also involve a certain amount of temporal inconsistency, i.e., latency. User interfaces that build on perfect spatial congruency and extremely low latency may be feasible in many situations, but this is certainly not always possible and overcoming the spatial limitations of the real world can be beneficial. We therefore prefer the term coherence, which implies that action and perception (or input and output) are not necessarily consistent, but that their interrelations follow a coherent logic that can be consistently perceived by all involved participants. The mapping of an input motion to a controlled object on a screen, for example, must not be isomorphous, but the relation between both should be clearly visible from every involved perspective. Following Gutwin and Greenberg's concept of workspace awareness [121], we furthermore limit the demand for coherence to the workspace, i.e., activity related aspects of the person and task spaces.

Several examples given above suggest a serious coordination overhead involved in cooperative action. Clear role assignments as suggested by the pattern of *specialization*, can alleviate this issue if the involved roles have minimal dependencies, but this is not always desired. Multi-user collaboration, for example, involves subsequent phases of tight and loose coupling between participants. These phases cannot be designed, but cooperating parties must be given the possibilities to organize their coupling spontaneously with respect to the relevance of subtasks and the potential benefits of different cooperation patterns.

Scott et al.'s observations of *emergent territoriality* indicates a promising design approach to this end [311]. They studied collocated collaboration on a shared physical tabletop. The interaction space was not initially subdivided. Through ongoing activities, however, dedicated zones for public exchange, private subtasks, and storage (subspaces of public and private territories) emerged. Four concrete implications for the design of tabletop interfaces were derived from these observations, that can be extended to cooperative user interfaces in general. According to Scott et al., group interaction benefits from 1. visibility and transparency of actions, 2. sufficient space, 3. correspondence of localized functionalities to the respective territories, and 4. support for casual grouping of items and tools [311]. The first point can be considered an aspect of workspace awareness as defined by Gutwin et al. [121], but the three other aspects emphasize the particular relevance of spatial structuring for group interaction.

During the tabletop collaboration studied by Scott et al., the characteristics of the environment developed from initial *equivalence* to habitual *specialization*. Similar processes can be observed when members of a group find their individual roles with respect to those of the others. The work of architects and designers largely consists of structuring space into zones for dedicated subtasks. In case of a large variety of potential applications and users with different capabilities, however, it seems advisable to enable more freedom or *equivalence*. The observation of *emergent territoriality* suggests a need to enable separation, but to leave the process of *specialization* to the users. Meaningful *specialization* promotes the autonomy of participants in the operation of self-contained subtasks with minimal dependencies.

6.3 Levels of Autonomy

Tightly coupled cooperation seems to be more readily attainable among different limbs, senses and organs of a single person. In social collaboration, independent organisms, each with their own brain and nervous system, need to coordinate their actions. This can be more difficult, as the information exchange is limited to the perception of external events. It is hard to imagine that the impressive performance of bimanual typewriting could be achieved by two typists sharing a single keyboard. Four-handed piano playing, however, indicates that this might be not so far-fetched. The key difference between both examples is the level of autonomy of the participants.

We have discussed in Chapter 4 that touch typing involves hierarchical motor control with words serving as the interface between conscious planning and low-level motor operation [210]. If two users were to share the same keyboard in order to type together, they would not be able to complete many words using only letters of their dedicated side, hence the low-level typing procedures would be severely impaired. Logan and Crump performed an experiment in which they asked typists to operate only one of both hands during the typing of a paragraph [209]. The performance of their participants fell from 80 words per minute with 6% errors to 14 words per minute with 33% errors. Apparently, as Logan and Crump put it, “the left hand doesn’t know what the right hand is doing”. As a two-user performance, the typists could learn that the missing letters are complemented by a collaborator, but this would require extensive training.

In the case of four-handed piano playing, both players are responsible for complete musical phrases that are meaningful per se. Similar to a duet with two pianos, the other player contributes to a more complex musical performance with complementary chords and melodies – in this case a few octaves lower or higher. The example illustrates that people in social cooperation must be provided with sufficient autonomy to maintain integral control over all aspects of self-contained subtasks. Interface support for the pattern of *territoriality* as described by Scott et al. can be helpful in that regard [311]. The two piano players in our example, divide the clavier into its left and right side and assume these as their private territories.

The allocation of keyboard regions to the left and the right hand during touch typing facilitates their cooperative performance, but apparently, such low-level subtask separation is not feasible for the cooperation of multiple people. Limbs belonging to the same person are naturally more tightly coupled with each other, but, also the looser coupling in collaborating groups can be advantageous. For a single person, it is difficult to perform asynchronous motor actions and almost impossible to focus consciously on multiple simultaneous processes. Group members, instead, can work almost completely autonomously. They may therefore better exploit the *division of labor* and higher-level *specialization* as cooperation patterns, since independent subtasks can be performed in parallel.

In fact, cooperating users often need a certain level of individual autonomy to perform better as a team. It has been shown, for example, that brainstorming sessions can be ineffective if the setting does not allow participants to work alone and take individual responsibility (see [299, pp.64-66]). Therefore, interfaces for multi-user cooperation should support fluent transitions between individual activities and varying levels of collaborative coupling. Loose coupling can increase the diversity of contributions, while tight coupling is required to achieve mutual agreement and convergence towards intermediate resolutions. Support for *territoriality* as an emergent social behavior seems to be a pragmatic, yet powerful, design principle in that regard.

The general pattern of territoriality is a long known behavioral pattern of humans and other animals. It describes the spatial structuring of an environment for interaction with others (see [80] for an introduction). In particular, private spaces are delimited to safely support activities of recreation and reproduction. The behavior is closely related to Hall's concept of proxemics [125]. Hall argued that humans try to maintain certain distances between each other that depend on their social relation. He defined four ranges of such interpersonal distances[1]: intimate, personal, social, and public. The habitual structuring of space, as described by territoriality, supports adherence to appropriate interpersonal distances (proxemics).

Edney noted that territoriality "has often been summarized as the defense of particular spaces by individuals and groups" [80, p.31]. He notes, however, that human territorial behavior is more complex. Humans "routinely entertain others on home ground without antagonism (as in visiting)" and human territory-related fighting almost solely occurs on the group, rather than the individual level. He suggests the more general criterion of "continuous association of person or persons with specific place" [80, p.33]. This definition also corresponds to the observations of Scott et al. [311].

---

[1] Ranges of interpersonal distances (with close and far phases): *intimate* (0 cm – 5 cm – 46 cm), *personal* (46 cm – 76 cm – 120 cm), *social* (1.2 m – 2.1 m – 3.7 m), *public* (3.7 m – 7.6 m – more) [125]. These measurements were derived from studies with adult US citizens and may vary among age groups and cultures, but they offer rough approximations for interpersonal distances in western culture.

6.4 Conclusion

Benefits of cooperative action have been demonstrated with multimodal, bilateral, as well as collaborative user interfaces. We have seen that interaction patterns that have been observed in one domain also occur in others. The reason for this broad applicability is the focus of these patterns on high-level cooperative task solving, rather than the low level operations of the involved actors. Based on our overview of cooperation patterns and requirements, we derived general requirements for cooperative user interfaces. The overarching design goal is the development of interactive systems that encourage multiple intertwined input actions while minimizing potential conflicts. We suggest to follow three high-level design principles: *workspace coherence*, *complementary capabilities*, and support for *emergent territoriality* (Figure 6.2). These do not prescribe a clear path towards effective cooperative user interfaces, but they put an emphasis on selected aspects, which we consider to be particularly relevant.

Mutual awareness seems to be the most fundamental requirement for any coordinated action. Such awareness can be achieved through intentional communication, e.g. explicit notifications, which generally induce the additional cognitive load of decoding and comprehension. We argue that *workspace coherence*, as described above, can minimize the necessity for such explicit communication. The awareness of others and their actions is an implicit result of continuous perception (feedthrough). In the case of bilateral and multimodal interactions of a single user, it is often sufficient that the perceived effects of all inputs support a consistent mental model. In the case of multi-user 3D applications, the appearance from different viewpoints must be considered. It has been shown that such workspace coherence can improve the fluency of multi user interaction, since it provides the necessary resources for effective negotiation [149]. *Workspace coherence* is the prerequisite for cooperation by *redundancy*, *transfer*, and *equivalence*.

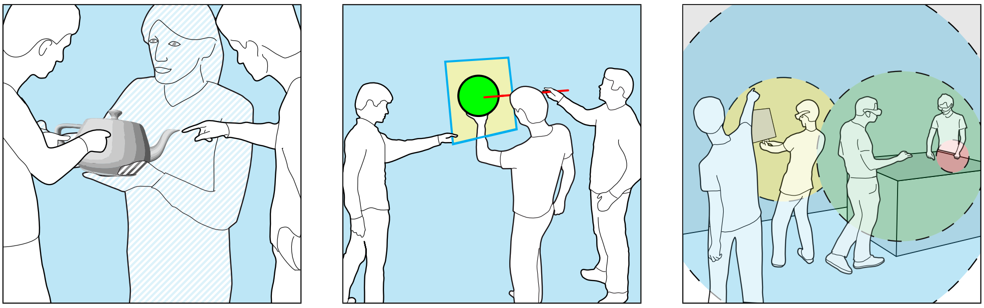
Workspace coherence supports awareness of oneself and others as well as consistent perception-action couplings in a joint workspace through spatial, temporal, and semantic relationships that can be readily perceived and traced by all cooperating parties. Joint perception, common sense, or mutual agreements can provide a suitable foundation for workspace coherence. When inconsistencies are unavoidable, the focus should be on a coherent representation of the workspace, i.e., areas of common relevance, joint interests, and current activities.

The full potential of cooperative action can only be exploited if the involved actors can support each other's actions, either in the sense of intensification (see Section 6.1.3 on *confirmation and intensification*) or through *complementarity*, including the patterns of *specialization* and *transfer*. *Confirmation and intensification* can be supported through support for parallel actions in a coherent workspace. *Complementary capabilities*, on the other hand, must be explicitly supported through the provision of specialized interaction capabilities and interfaces for their combination.

> *Complementarity capabilities* encourage cooperation and reduce potential interaction conflicts by providing multiple parallel options for user engagement and extended interface functionalities based on their combination.

Moreover, cooperation processes between multiple users generally involve subsequent phases of loose and tight coupling. Complementary actions are not always performed at the same location or on the same object. Collaborators frequently deviate from a shared group space to private exploration and prepare information or items for later exchange. The emergence of distinct interaction territories is often task-related and can also be observed during single user interaction. Consider for example, the multiple open windows on a desktop computer and the many stacks of paper related to different activities that often surround the keyboard and the display. User interfaces for cooperation should support such multitasking and *emergent territoriality*, e.g., through the immediate availability of multiple interaction areas and support for continuous spatial restructuring.

> *Support for Emergent territoriality* facilitates frequent transitioning between loosely and tightly coupled cooperation through user-defined partitioning for different levels of privacy and group exchange.



(a) Coherence (b) Complementarity (c) Territoriality

Figure 6.2: *We suggest three core topics for the design of cooperative user interfaces: workspace coherence, complementarity capabilities, and support for emergent territoriality.*

Part II

Approach and Experiments

Chapter 7

The Design and Evaluation of Cooperative User Interfaces

The paradigm of cooperative user interfaces implies support for multiple concurrent actions and their synergistic combinations. The number of such combinations grows exponentially with the number of involved actors and their capabilities. Consider piano keys as a very basic example. A single finger can strike only one key at a time. With two fingers one can choose between two notes to be played individually, simultaneously, or not. The choice is between four different sound effects. If we consider their cooperation over time, the two simultaneously available notes can also be combined in direct succession, e.g., playing a trill. A full hand can do the same with five notes, resulting in 32 different sound effects that are available in a single moment. Variations in timing and loudness enable infinite expressive possibilities – but not all. Two hands can certainly create more complex sounds and two players with four hands even more so. The players may also contribute with their feet by operating the pedals of the piano. Such a cooperative action must be well coordinated to create complex music instead of a cacophony.

The concept of cooperative user interfaces suggests that any user action can be performed in combination with others. Ideally, multiple users can simultaneously induce motion input from all their limbs and additional modalities like gaze and speech to operate a diversity of tools in synergistic cooperation. The quality of different interface designs supporting such a rich interaction, can hardly be compared. A comprehensive evaluation would require the usability analysis of each individual interface parameter and all possible combinations of use. Several effects of individual parameters are known, others can be experimentally explored, but a comprehensive analysis of all possible combinations and their interaction effects seems unattainable.

The design of such complex systems is often an under-constrained problem with many well suited configurations but no global optimum. It cannot be conclusively solved through modeling and optimization. Instead, the effects of design decisions on the users, their experience, task performance, and behavior must be deduced a priori from experiences with similar situations. The validity of these estimates can be only partially tested through post-hoc evaluation of concrete implementations.

Rittel and Webber described the prevalence of such "wicked problems" in planning and specifically in policy making. They also specified a number of common characteristics that stem from the ill-defined problem space and potentially high stakes (see [284]): Wicked problems cannot be formally defined; Each wicked problem is interrelated with several others; An ultimate root problem cannot be identified; As a result, there is no optimal solution, but an infinite amount of potential (temporary) resolutions, which depend primarily on the world view of the analyst, planner, or designer; Every wicked problem is unique; The transferability of existing solutions to similar situations is uncertain, as any distinguishing aspect can be overruling apparent similarities; The effects of implemented configurations cannot be comprehensively evaluated since relevant side effects may be obscured; Moreover, Rittel and Webber note that trial and error is not an option if high stakes are involved (e.g. in public policy), since every attempt may have severe consequences.

Fortunately, the stakes for the exploration of novel user interfaces are much lower than those involved in public policy. A new design must not be adopted by everyone, instead, it extends the users' choice among a diversity of interface options. Gaver argues that an "... endless string of design examples is precisely at the core of how design research should operate ..." [104, p.938]. He continues that there is no right or wrong in design, but many simultaneously existing worlds, each proper in its own right. Fallman previously highlighted the pivotal role of design in innovation-driven HCI research [82]. Building prototypes in order to explore the effects of new interaction concepts and ideas is essentially a design process. Additionally, he argued, the knowledge generated in scientific studies directly influences the design of novel products in industry. He distinguished between design-oriented research, where the goal is the creation of knowledge, and research-oriented design, which aims for the creation of new products based on better informed design decisions. The latter case is the well established application of proven scientific knowledge to the design of artifacts. The first case, however, raises questions on scientific methodology.

7.1 Research Through Design

The concept of design-oriented research, or research through design, emphasizes the impact of design processes on the generation of scientific knowledge. Research tools and prototypes must be first designed and implemented to enable new understandings. Fallman indicated that the involved design processes may not live up to the methodological rigor expected from a scientific discipline. He refers to descriptions of design as an iterative process that is inherently non-linear. Design thinking, in his account, takes the form of sketching and prototyping, which cannot be considered rational, transparent methods with clearly separable phases of analysis, synthesis, and evaluation. Instead, problem setting and problem solving are "intertwined in the activity of design, an inseparable pair only unfolded through the design dialogue" [82, p.230]. The holistic approach of design does not comply with the concept of decomposition commonly applied in science and engineering. According to Fallman, design should be considered a cultural tradition of thought and action rather than an academic discipline somewhere between art and science. Ultimately, the goal of design is not theory but the production of tangible artifacts.

Zimmerman et al. largely agreed with the argument of Fallman, but they further suggested that design thinking can be structured by the phases of investigation, ideation, and iteration [384]. Compared to the engineering approach of analysis, synthesis, and evaluation each individual phase fosters divergence instead of convergence. Zimmerman et al. suggested that investigation in the design process refers to the assumption of multiple perspectives at a situation rather than the decomposition of a problem; Ideation signifies the suggestion of multiple alternative options instead of the synthesis of one optimal solution; Iteration aims for improvements and adaptation to changing settings rather than the validation of a particular solution. Engineering also starts with a broad exploration of the design space, but its goal is an optimal solution, whereas design aims for a diversity of options and styles.

Zimmerman et al. suggested four criteria for the evaluation of design research focusing on the process, the extensibility, the inventive step, and the relevance of proposed designs [384]. Accountability of a design process does not imply its replicability, but nevertheless, the reasons for selected methods and the applied rigor should be clearly conveyed. Moreover, the documentation of the design rationale and its implementation should support extensibility to other situations and setups. Design aims at the creation of new artifacts, hence the level of inventive ingenuity should be clarified through an extensive review of related work. Validation, as generally required for new scientific theories and models, is not always applicable to new designs. Therefore, Zimmerman et al. propose that design research should articulate the goal of an invention and its relevance.

On a more general level, Gaver recently discussed design theory and its academic treatment [104]. Design theory is often expressed in concepts, principles, and guidelines, which suggest, but do not guarantee, potential benefits. One can always find cases where adherence to design principles, did not lead to the desired effects. Although falsification is the established method of disproving scientific theories [271], such a failure of principles in a particular case, does not disprove a design theory in general. Gaver argued that design theories do not describe "*what is* " but "*what might be*". Therefore, they are not falsifiable and cannot be evaluated by commonly upheld scientific methods [104].

Gaver offers Lakatos's notion of research programs as an alternative approach to understanding the role of theory in design. Such a research program consists of a conceptual core that is protected by auxiliary hypotheses and offers guidance in problem solving [192, in [104]]. Correspondingly, Gaver argues that research through design contributes "theories that are provisional, contingent, and aspirational" [104, p.937]. Their meaning and values can be best expressed through a number of examples. Gaver thus suggested the concept of annotated portfolios. The primary role of theory in this account is the annotation of product portfolios that contain examples of a representative area of the design space. In this sense, Gaver considers annotated portfolios in opposition to design patterns. "They are not intended to abstract regularities from repeated attempts to design for the same domains. Instead, they maintain the particularity of individual examples, while articulating the ideas and issues that join and differentiate them." [104, p.945]

7.2 Research Program and Working Hypotheses

Our research program is focused on the design of user interfaces that support and encourage cooperation. The paradigm of cooperative action had been proposed earlier for special cases of bimanual object manipulation [138], multi-user collaboration [46, 234, 269, 293], or multimodal interfaces [61]. Chapter 5 discussed several prior interaction systems and their support for cooperative action. This thesis proposes consideration of cooperation as an overarching interaction principle that can be consistently applied to different types of interactive computer systems. Interfaces that support cooperative action enable higher versatility, flexibility, and expressiveness for single users. At the same time, they encourage the involvement of multiple users, through the redistribution of subtasks.

From a review of literature in the fields of human motor control, interpersonal coordination, and HCI, we deduced that the design of cooperative user interfaces should evolve around three main themes: *workspace coherence*, *input complementarity*, and *emergent territoriality*. We have discussed the relevance of these aspects for single-user and multi-user systems. However, such high-level design principles cannot guarantee successful results. One can easily come up with inappropriate designs despite adherence to these principles, e.g., if the application case or the user requirements are not fully understood or appreciated. In particular, the choice of interface functionalities, their ergonomics, and their accessibility depend on the application content, the interaction environment, as well as the goals and capabilities of the involved users. More elaborate heuristics, like those of Baker et al. for groupware usability [19], offer more detailed guidance, but, they neither provide readily applicable solutions. The suggested principles of cooperative user interface are deliberate abstractions of core aspects that can be applied to interface design more generally.

The research presented in this thesis is based on the following working hypotheses:

H1 Interface support for cooperative action is beneficial for single users and, at the same time, facilitates the collaboration of groups.

H2 The combination of concurrent user actions enables improved workload balancing, higher interaction fluency, more flexibility, and extended interface functionalities.

H3 The design of cooperative user interfaces can be successfully informed by the high-level design principles *workspace coherence*, *emergent territoriality*, and *complementary capabilities*.

These high-level hypotheses describe a fuzzy relationship between the design paradigm of cooperative user interfaces and potential usability benefits. They may not be proven to be completely true or false. As an alternative form of support, Gaver's concept of an annotated portfolio suggests a compilation of representative examples. For these examples, a detailed account of the invention, its development process, relevance, and extensibility as requested by Zimmerman et al. can be considered an extended annotation (see above). Where possible, also a formal evaluation of concrete goals and expectations can be added. Moreover, a qualitative analysis of coordination behavior can extend established evaluation measures of user performance and satisfaction. The experimental results can only account for a particular configuration of parameters, but if the configuration is a relevant one, they can increase the plausibility of the provided design annotations.

The chapters of Part II report in detail on three concrete developments of user interfaces for cooperation. All three have been published as mentioned in the list of publications on page xxv. The three reports comply to a certain degree with the above discussed evaluation criteria, but each with a different emphasis.

The first example takes a previously proposed 3D input device [98] and explores its benefits in a bimanual interaction task. This work emphasizes the relevance of ecological validity in the evaluation of interface designs. In contrast to common performance studies with single-handed aimed movement tasks, the expected benefits of certain interface characteristics are less pronounced when tested in a holistic sequence of cooperative actions. Small but significant improvements of general control accuracy could be obtained and a post-hoc task decomposition revealed the actual effect size for the affected subtask only.

The second example describes a novel multitouch interaction technique for mode switching. The design of this technique was based on a theoretical model of bimanual motor control and the formal specification was based on a detailed analysis of temporal behavioral patterns. Here, the focus is on the invention of the technique and the process of its development. The results of a formal comparison against an established mode switching technique revealed performance advantages that can be attributed to differences in cognitive processing.

The third report features a novel 3D projection system that provides up to six users individually with stereoscopic image pairs. We discuss the relevance of such a system for 3D visualization applications, describe its functional principles, and detail our design rationales. Moreover, we describe the challenge of group navigation in shared virtual environments and suggest novel interaction techniques as a resolution. A formal user study revealed benefits of the system for cooperative visual search as well as the usability of our novel group navigation techniques.

These three examples have been chosen because they reflect the diversity of research and design approaches that can be pursued towards cooperative user interfaces. The combination of multiple existing interface technologies enables multiple input streams, novel interaction techniques can increase the expressiveness of such parallel input actions, and novel technologies create a whole new context for the development of cooperative user interfaces. These three examples form the basis of a portfolio of cooperative user interfaces which is extended in Part III of this thesis with additional application examples of the derived design principles.

Chapter 8

Bimanual Cooperation with Desktop 3D Input Devices

This chapter reports on joint work with Jan Hochstrate, André Kunert, and Bernd Fröhlich at Bauhaus-Universtität Weimar. It has been presented at IEEE 3D User Interfaces 2009 and was published in the conference proceedings under the title:

“The influence of input device characteristics on spatial perception in desktop-based 3D applications ”

Abstract

In desktop applications 3D input devices are mostly operated by the non-dominant hand to control 3D viewpoint navigation, while selection and geometry manipulations are handled by the dominant hand using the regular 2D mouse. This asymmetric bi-manual interface is an alternative to commonly used keyboard and mouse input, where the non-dominant hand assists the dominant hand with keystroke input to toggle modes. Our first study compared the keyboard and mouse interface to bi-manual interfaces using the 3D input devices SpaceTraveller and Globefish in a coarse spatial orientation task requiring egocentric and exocentric viewpoint navigation. The different interface configurations performed similarly with respect to task completion times, but the bi-manual techniques resulted in significantly less errors. This result is likely to be due to better workload balancing between the two hands

allowing the user to focus on a single task for each hand. Our second study focused on a bi-manual 3D point selection task, which required the selection of small targets and good depth perception. The Globefish interface employing position control for rotations performed significantly better than the SpaceTraveller interface for this task.

8.1 Introduction

The majority of 3D graphics applications are still desktop-based, which is largely based on ergonomic reasons. The physical support for the operating hand on the desktop surface efficiently reduces fatigue. While there is a broad variety of interfaces for immersive virtual environments, desktop-based 3D applications are mostly managed with the familiar mouse and keyboard set-up. The operation of such 3D applications requires a lot of mode changes if only a 2D mouse and a keyboard are used. In this case, the non-dominant hand assists the dominant hand with keystroke input to toggle modes. This approach has two major drawbacks: The workload distribution between both hands is very unbalanced and integral 3D manipulations need to be separated into a sequence of 2D actions. Prior work of Jacob et al. [157] and Hinckley et al. [141] indicates that the second issue can affect performance.

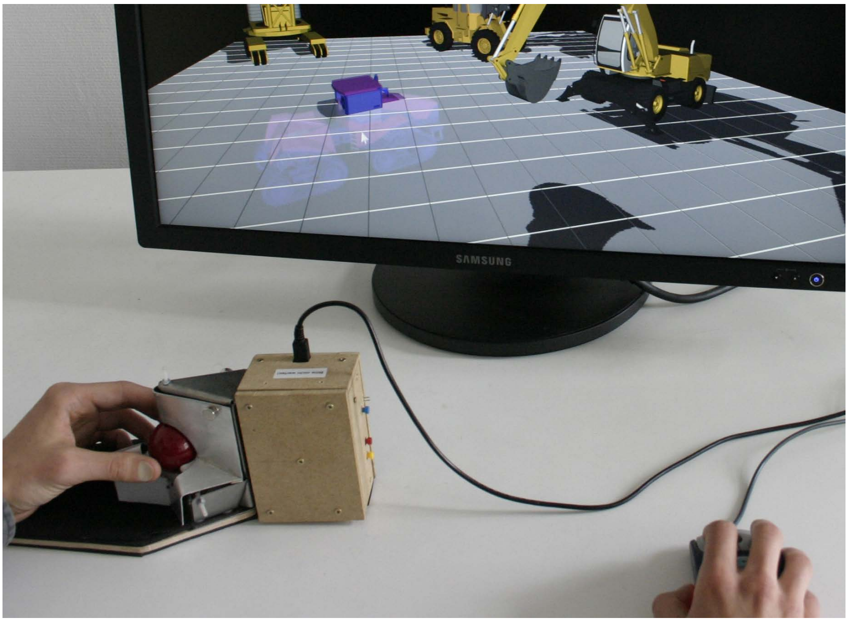


Figure 8.1: *The Globefish input device for CAD and DCC applications*

The SpaceMouse™ (with its smaller descendants SpaceTraveller™ and SpaceNavigator)™ is one of the few specialized 3D input devices that has gained respectable acceptance among users. The design enables the integral operation of 3D translation and rotation via elastic rate control. The 3D input device Globefish [98] separates translational and rotational input by its hardware design (Figure 8.1). Such 3D input devices are mostly operated by the non-dominant hand to control 3D viewpoint navigation, while selection and geometrical manipulations are handled by the dominant hand using the regular 2D mouse pointer. However, there is no scientific evidence if and why such an asymmetric bi-manual interface configuration is a good choice.

In a first step we analyzed the interaction requirements of desktop 3D applications. Based on our observations we implemented two scenarios and evaluated them by user studies. The first study compared regular keyboard and mouse input to two-handed input using a 3D input device and a mouse. The task focused on coarse spatial orientation in egocentric and exocentric viewpoint navigation. We constrained the required degrees of freedom such that they could be directly provided by the 2D mouse. Our results indicate that the input device configuration does not have much influence on the time efficiency in coarse spatial orientation tasks. However, bi-manual techniques resulted in significantly less errors. We argue that this observation is due to better workload balancing between the two hands allowing the user to focus on a single task for each hand. Our second study compared the two 3D device concepts SpaceTraveller and Globefish in a bi-manual 3D point selection task. The task required high selection accuracy and good depth perception, which had to be achieved through 6-DOF view point navigation. The results of this study revealed significant benefits for the position-controlled rotational input provided by the Globefish device.

8.2 Related Work

Hinckley et al. [141] demonstrated that using appropriate input devices to control 3D rotations can be very beneficial. Compared to mouse-driven techniques like the Virtual Sphere [59] and Arcball [317], the average task completion times were reduced by up to 36% with orientation-tracked handheld devices in a 3D orientation alignment task. Balakrishnan et al. [21] instead described a number of advantages of the mouse-based input even for 3D interaction. From this analysis and the sustaining demand for more integrally provided degrees of freedom, they derived the concept of the Rocking Mouse. The device is a variation of the mouse that enables simultaneous control of four degrees of freedom instead of only two. A comparison to basic mouse functionality in a 3D positioning task demonstrated important advantages for the device. The results of both studies support the theory of Jacob et al. [157] that simultaneously required degrees of freedom of a task should also be integrally incorporated into appropriate input devices.

Devices like the SpaceMouse provide six integral degrees of freedom for the integral operation of 3D rotation and translation. The device employs rate control for both tasks, which allows moving virtual objects around an unlimited workspace through minimal deviations of the input handle. While this is necessary to control 3D translations without lifting the arm from the supporting desk, rotational input could also be provided with position control as in the case of the Globefish device. A comparative user study revealed the superiority of this approach for 3D object manipulation over the fully integrated 6-DOF design of the SpaceMouse, providing only elastic rate control [98]. The separation of rotational and translational input seemed to be beneficial for the chosen 3D docking task which helps to explain the results, but several other studies also indicate advantages of position control over rate control.

Shumin Zhai showed that both techniques may perform equivalently, if used with the appropriate devices, but he also found higher training requirements for rate control [376, 378]. In an experiment using a one-dimensional scrolling task, Hinckley et al. [134] observed advantages for position control to operate short range movements and that this range can essentially be extended with software techniques such as pointer acceleration. Kunert et al. [190] validated this interaction of distance and technique in a 6-DOF manipulation task. Since rotations are cyclic, they rarely exceed the range where position control is beneficial. Thus, it is not surprising that Kim et al. [172] observed position-controlled object rotations with an isotonic 3D trackball to be more efficient than operating the same task with the rate-controlled SpaceMouse input.

In practice such desktop-based 3D devices are operated by the non-dominant hand to control the view at a scene, while more frequent pointer interaction is assigned to the dominant hand. The efficiency of such a workload distribution in similar contexts has often been demonstrated [22, 115, 151, 167, 204]. Previous studies on specialized 3D interaction devices in desktop environments only analyzed performance of operations of the dominant hand. We aim to close this gap, by analyzing input device performance with the non-dominant hand in compound bimanual tasks as they are typically found in common 3D applications like games, CAD and DCC.

8.3 Desktop-Based 3D Interaction

3D games as well as 3D modeling in engineering and art are highly successful 3D graphics applications. Most often they are operated by devices known from 2D applications, with the mouse as the primary input tool. On closer inspection of the requirements of those applications we found that this adaptation of an existing infrastructure is not inappropriate. The major interaction task performed is the selection of trigger buttons, tools and objects. Using image plane selection techniques, even selection in 3D space can essentially be simplified to controlling only two degrees of freedom, for which a regular 2D mouse is perfectly suited. Since the task is the most frequent one and requires high precision, it is naturally assigned to the dominant hand [115].

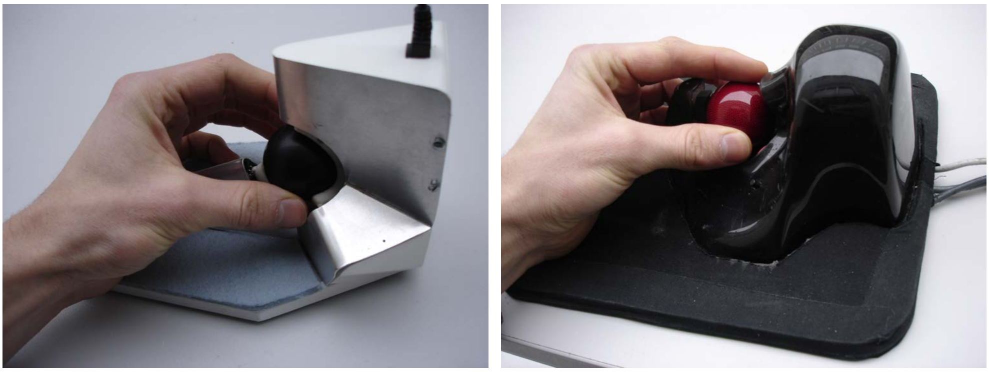
To extend the applicability of the device, the non-dominant hand assists with keystroke input to toggle modes and tool selection shortcuts. If navigation and manipulation are required however, interaction becomes more difficult. All adjustments of fine-grained parameters are necessarily assigned to the mouse device. Many efficient techniques have been developed to map 3D interaction through 2D input [59, 317, 319], but this approach also increases the workload of the dominant hand and thus its physiological fatigue. To cope with this problem one can assign more sophisticated input devices to the non-dominant hand. This can be another pointing device [375] or a specialized 3D navigation and manipulation device like the SpaceMouse. The latter is more common and appreciated by many professional users. In our experiments we analyze the impact of controller design for such two-handed interaction with respect to the specific requirements of egocentric viewpoint navigation and exocentric object examination.

8.4 Desktop-Based 3D Input Devices

The Spacemouse is a desktop-based input device that enables users to control motion direction and velocity of 3D rotations and translations with small deviations of an elastically suspended handle. Users often comment on its comfortable operation but also about difficulties to work accurately with the device. Linear motion can hardly be induced without rotational input and vice versa. Furthermore, rate controlled input is not as simply reversible like position controlled input, which was pointed out by Balakrishnan et al. [21]. The device is offered in a number of different designs that vary in the size of the handle as well as the number of additional input buttons. In the described experiments we used the SpaceTraveller with a handle diameter of 45 mm (Figure 8.2).

Figure 8.2: *The Spacetraveller 3D input device*

The Globefish is a novel sensor concept for desktop based 3D interaction. For our experiments we used prototypical implementations of the device. It consists of a 3DOF trackball which is elastically suspended within a surrounding frame. The trackball, suitable for being precisely held from two opposing sides, can be slightly moved in all spatial directions to induce translational input. Rotational input is generated by simply rotating the sphere, while translations are induced against an elastic counterforce similar to the Spacemouse device. For the experiments we used two different prototypes of the device. Both consist of a 40 mm Trackball and provide elastic counterforces similar to those of the SpaceTraveller. The design of both emphasizes rotational input around the vertical and the lateral axis as well as translational input in depth direction, since those are the most frequently required degrees of freedom. Both offer ergonomics that afford a comfortable hand posture similar to writing with a pen. The device prototypes differ in resolution and positioning of the rotation sensors and also they differ slightly regarding their ergonomics. While the more recent prototype used for the first study on coarse spatial orientation exposes the trackball towards the operating hand (Figure 8.3 a), the trackball of the antecedent prototype is laid open at its top (Figure 8.3 b). Though we believe that the recent design is slightly better, we also tested with the previous prototype due to technical issues. Regarding the results of the experiments we argue that the differences between both are negligible.



(a) Globfish prototype I

(b) Globefish prototype II

Figure 8.3: The two Globefish device prototypes used in the studies

8.5 Spatial Orientation Study

3D graphics applications allow for viewpoint motion to handle occlusion problems and to make larger workspaces accessible. Typically, it is not possible to position the viewpoint as such that all relevant objects in the scene are visible. Thus, egocentric viewpoint rotation as well as exocentric rotations around an object of interest are frequently required. The difference between both is the rotation pivot. The viewpoint itself defines the center of rotation within the egocentric case, enabling the user to look around in the virtual environment. In the exocentric case the pivot refers to an external object in order to encircle it. Accordingly, our study included alternating subtasks of exocentric and egocentric viewpoint motion.

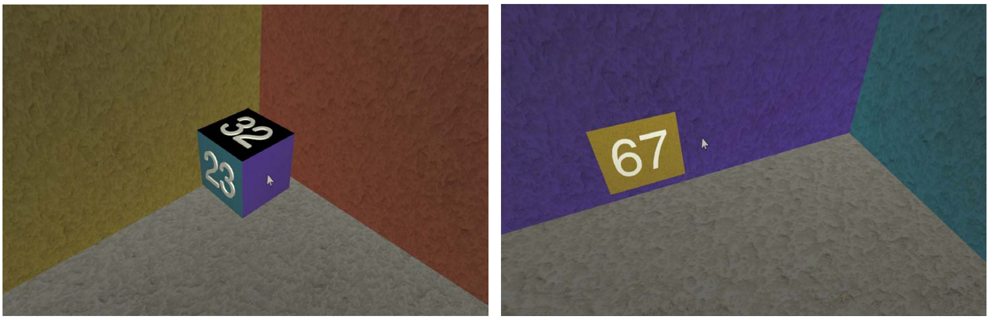
8.5.1 Task

We have chosen a very simple environment for our task. It consisted of a cubic room and a much smaller cube inside that room. The faces of both were displayed in six different colors. The small cube was aligned with the cubic environment and placed at its center (Figure 8.4 a). For visual orientation, we assigned black to the top, a light gray to the bottom and saturated colors to the remaining four faces of the room and the cube.

The task started in exocentric mode with the viewpoint halfway between the borders of the cubic room and the smaller object at its center. The viewpoint was automatically oriented towards the center and could only be moved in two dimensions (head and pitch) encircling the smaller cube. The exocentric navigation technique is similar to rotating an object in front of the view (object manipulation) except two important differences: 1. Not a specific object is manipulated, but the whole environment remains consistent while the viewpoint is rotated around an external pivot. 2. The environment in front of the view therefore appears to rotate in the opposite direction, as the viewpoint is moved by the user's input. In practice, both methods can be found to support object examination. We decided to conduct the test with exocentric viewpoint navigation instead of an object manipulation technique to comply with the camera-in-hand metaphor consistently in both orientation subtasks.

Our daily experience with spatial orientation in the real world is largely influenced by gravity effects. Humans are therefore most familiar with wayfinding in environments that are more or less horizontally aligned. To prevent loops that would easily lead to disorientation, we limited the arc motion around the lateral axis to $+/-60\,^{\circ}$. The participants were instructed to unveil numbers behind each of the cube's six faces and memorize the highest one. Mouse pointer input was controlled with the dominant hand to enable the selection of the target faces, which triggered randomized numbers between 10 and 99 to appear on the selected surface patch. The exocentric subtask could then be finished through double clicking on the face with the highest number displayed. Thereafter, an animated transition moved the viewpoint to the center of the scene, where the small cube was previously located.

The second phase of the task, starting after the automated viewpoint transition to the center of the scene required the user to control egocentric viewpoint rotations. Only rotations around the vertical (head) and the lateral (pitch) axis were enabled. To



(a) The exocentric rotation sutask (b) The egocentric rotation subtask

Figure 8.4: *During the exocentric subtask (a), the viewpoint had to be turned around a cube in the center of the scene to explore the object from each side and select its six faces in order to unveil a randomized number. During the egocentric subtask (b), the viewpoint was located at the center of the cubic environment. Quadratic target patches could be found on the surrounding four walls that had to be selected in order to unveil a randomized number. In both cases, the highest number had to found and confirmed through selection with the mouse pointer.*

avoid disorientation, rotations around the lateral axis were again limited to $+/-60\,^\circ$. Here, the users had to search for smaller squares located at the edges of the four surrounding vertical walls. In contrast to the exocentric subtask, we excluded targets at the top and the bottom face of the surrounding cube, since finding these would have caused more difficulties. To involve also rotations around the lateral axis, the target squares were situated at different heights. Two opposing walls showed the squares at their lower edge and the other two at their upper edge. These squares were sized as such they appeared in a comparable size on the image plane as the surface patches of the target cube in the exocentric subtask. To ensure good visual contrast, target squares had the same color as the opposing wall segment of the surrounding room (Figure 8.4 b). Again, participants were asked to unveil randomized numbers (ranging from 10 to 99) behind these targets through mouse pointer selection. In the egocentric condition four targets had to be unveiled before the users could decide on the highest displayed number among them and then select it with the mouse pointer. After finishing the task by double clicking with the mouse pointer on the respective highest number, an animated transition moved the viewpoint again to the starting position for the next exocentric subtask.

For both subtasks, a distribution algorithm assured that the required rotation angles ($0\,^\circ$, $90\,^\circ$, $180\,^\circ$, $270\,^\circ$) to move from the last unveiled target square to the one with the highest number were randomized and that they added up to the same amount of required motion for each block of trials.

8.5.2 Apparatus

The study was conducted on a desktop workplace with the user seated in front of a table providing a support for the input devices and the graphics display (20″ wide-angle LCD with a resolution of 1680x1050 px). The test application was running at 60 Hz. In all conditions, the user's dominant hand manipulated the virtual pointer with a regular mouse device. A linear transfer function with a control-display gain of 10 was assigned to the mouse pointer. The large targets did not require much accuracy and thus allowed for such a high-gain transfer function. Three input device configurations were used for viewpoint control.

In two device conditions the non-dominant hand controlled a 3D input device: either the SpaceTraveller or the Globefish device. Since the navigation subtask also involved only the control of 2DOF, we included a basic mouse/keyboard condition in the tests. In this condition, pointer manipulation as well as viewpoint control were assigned to the mouse (using the same linear transfer function) and operated by the dominant hand. This approach obviously required mode changes. The participants had to hold the CTRL-button on the keyboard to trigger viewpoint navigation mode. The graphical pointer remained visible and also button clicks were enabled in this mode. It was possible to accomplish the whole task while remaining in viewpoint navigation mode. Otherwise, frequently required mode switching through keystrokes would interrupt the workflow and users could hardly perform on par in the mouse condition.

The elastic SpaceTraveller was used with rate-control and a non-linear transfer function for high precision at low velocities, while still enabling rapid movements with larger deviation of the device's handle. Both isotonic devices, the 2D mouse and the 3D trackball of the Globefish, were used with position control and linear transfer functions. For the 3D trackball, we implemented an isomorphous mapping since we assumed that users may benefit from a congruent relation between the amount of input motion and the resulting rotation on the screen.

8.5.3 Hypothesis

In a comparative study on the impact of rate control and position control on document scrolling performance, Hinckley et al. [134] found advantages for position control. They further demonstrated that such scrolling tasks can be modeled with Fitts's Law. Andersen argued, that this finding is only true, if the target position is known beforehand. In contrast, he found a linear relationship between the time required for scrolling a document and the distance to be covered, if the target distance is unknown [7]. He assumed that in this case, rate control might be better suited, since it facilitates motion with a constant velocity.

In our case, only short moves had to be accomplished and since the environment was very simple, we assumed that users would be quite conscious about the required amount of rotation to find the next target for selection. Position control provides proprioceptive cues on the amount of motion input induced. The counterforce of elastic devices for rate control on the other hand provides haptic information on the motion velocity, which is more indirect regarding the goal to reach a certain position. Thus we hypothesized that position control with both isotonic input devices (Globefish and mouse) would result in better performance compared to elastic rate control with the SpaceTraveller, because users can benefit from proprioception. We expected the mouse to show the best performance, since it is well suited for this specific task and most users are really proficient with the device. We also assumed that additionally operating a task specific input device with the non-dominant hand can increase time efficiency and/or accuracy due to the distribution of workload.

8.5.4 Design and Procedure

After a short training session to accommodate to the task environment, three blocks of 12 trials were recorded with one device condition. As described above, each trial included an exocentric and an egocentric subtask. Short breaks between the subsequent blocks helped to minimize fatigue. Thereafter, the same procedure was applied to the other two device conditions. The order of devices was fully permuted between six independent groups. After completing the test with the three device configurations, the participants were asked to rate the three tested input devices on a five point Likert scale. Overall, the experiment lasted about one hour.

8.5.5 Participants

Twenty-four volunteers, aged between 19 and 34 years, took part in the first study. All of them were students from various disciplines including humanities, engineering and fine arts. Eleven among them were male and thirteen were female. All except one were right handed. The mouse was always operated with the dominant hand, while the devices at test were operated with the non-dominant one, hence, we adapted the input device mappings to the left handed person. Half of the participants reported that they were familiar with 3D applications, ten reported to have had only marginal experience and two of them were complete beginners. Fifteen had no experience with using devices like the SpaceMouse, the other nine already had used such 3D input devices. Twenty-four volunteers, aged between 19 and 34 years, took part in the first study. All of them were students from various disciplines including humanities, engineering and fine arts. Eleven among them were male and thirteen were female. All except one were right handed. The mouse was always operated with the dominant hand, while the devices at test were operated with the non-dominant

one, hence, we adapted the input device mappings to the left handed person. Half of the participants reported that they were familiar with 3D applications, ten reported to have had only marginal experience and two of them were complete beginners. Fifteen had no experience with using devices like the SpaceMouse, the other nine already had used such 3D input devices.

8.5.6 Results and Discussion

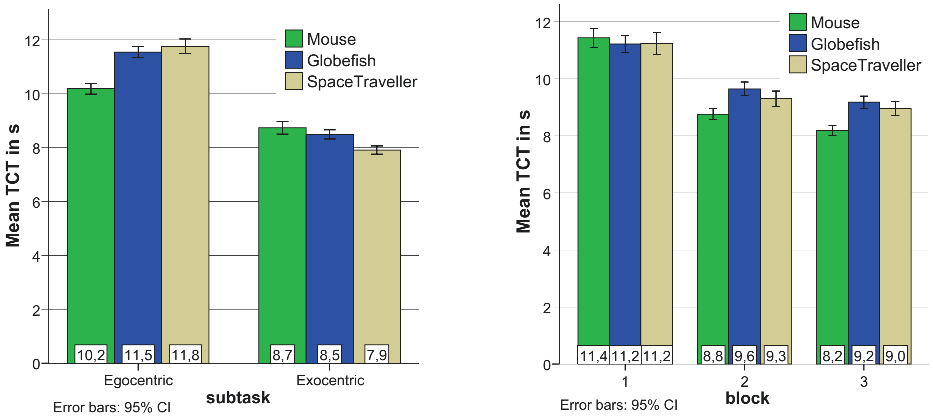
Having the device conditions fully permuted between six independent groups, we first tested on possible differences between groups, but did not find significant effects. Thus, all data was collapsed and entered into a 3 (devices) x 2 (subtasks) x 3 (blocks) analysis of variance, using Bonferroni adjustment for α in post-hoc comparisons.

Regarding task completion times, we found significant effects for the factors *subtask* ($F_{1,23} = 187.518$, $p < .001$) and *block* ($F_{2,46} = 211.344$, $p < .001$). Since both subtasks were quite different regarding their operation, it was expected, that the respective task completion times would also differ significantly (8.38 s in the exocentric and 11.17 s in the egocentric condition). The exocentric subtask was much easier, since the visual focus was fixed on the target object that defined the center of rotation. Thus, it could not be visually lost. Continuous learning can be observed over the three blocks. Post-hoc comparisons showed that the improvements between subsequent blocks are all significant (all $p < .001$). Average task completion times decreased from (11.30 s) over (9.24 s) to finally (8.78 s).

Device produced no main effect. This indicates, that all three tested device conditions are comparable in terms of time efficiency. In detail though we found the *device* condition to interact significantly with *block* ($F_{4,92} = 3.61$, $p < .01$) and even stronger with *subtask* ($F_{2,46} = 27.81$, $p < .001$).

The interaction between device and block (Figure 8.5 b) seems to result from a different learning behavior for the mouse. Both input conditions involving motion input from the non-dominant hand expose smaller and more consistent learning effects than the mouse condition. For the mouse condition we observed the strongest performance gain from the first to the second block. In average over both *subtask* conditions the mouse showed worst performance during the first block (11.44 s), but showed the best performance in the following two (8.76 s and 8.19 s). This indicates that with the mouse participants required learning only for the task, since they were already highly proficient with the device. Post hoc comparisons on the interaction of device and block using the Tukey-Test revealed one significant effect, namely an advantage of the mouse over the Globefish during the last block.

A Tukey test on the interaction of device and subtask (Figure 8.5 a) revealed a significant advantage of the mouse to both other devices in the egocentric viewpoint subtask (both $p < .01$), but no further significant effects. This subtask dependency can also be observed in subjective ratings (from 1=best to 5=worst). While users preferred the mouse for egocentric rotations (Mouse: 2.20, SpaceTraveller: 2.41, Globefish: 2.54), they favored the 3D input device conditions for the exocentric subtask (SpaceTraveller: 1.54, Globefish: 2.08, mouse: 2.20). Though we did not find a significant interaction between device, block and subtask, we observed stronger training effects for the mouse in the exocentric condition. Some users also reported of having been confused from using the device with an exocentric navigation technique.



(a) Sorted by *device* and *subtask* (b) Sorted by *device* and *block*

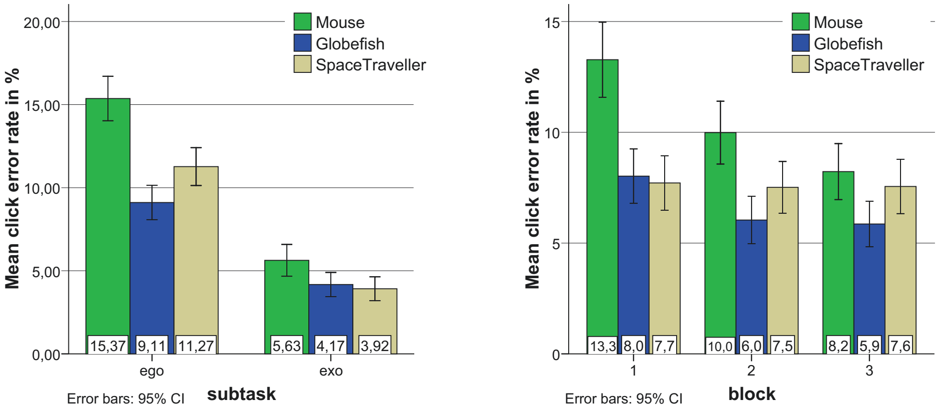
Figure 8.5: *Mean task completion in seconds*

We also recorded click errors per trial. All kinds of unnecessary button events (repeated selection, empty space selection, selection of a wrong target face) were counted as such. The number of click errors was divided by the number of necessary selections to compute the relative click error rate. An analysis of variance (using Bonferroni adjustment for α) revealed significant main effects for *device* ($F_{2,36} = 19.37$, $p < .001$), *block* ($F_{2,36} = 14.73$, $p < .001$) and *subtask* ($F_{1,18} = 93.71$, $p < .001$) as well as interaction effects of *device* with *block* ($F_{4,72} = 4.68$, $p < .05$) and *device* with *subtask* ($F_{2,36} = 5.75$, $p < .05$). The egocentric subtask resulted in a significant higher errors rate (11.92%) compared to the exocentric (4.57%) subtask. Accurate pointing was easier in the latter condition since the selection target remained in the center of the screen and thus only little movements of the mouse pointer were required.

Block effects revealed a significant improvement from the first (9.67%) to the following blocks (7.85% and 7.21%). This is an expected learning effect.

More interesting are the differences of devices. While we did not find many differences between the three tested *device* conditions regarding time efficiency, both two-handed techniques provided higher accuracy. Using only the mouse device to the dominant hand with mode changes resulted in 10.5% click error rate. Employing an additional 3D input device to the non-dominant hand resulted in a significant lower click error rate: 7.59% for the SpaceTraveller and 6.64% for the Globefish (both $p < .01$).

A Tukey-test on the interaction of *device* with *subtask* revealed that only in the egocentric condition the mouse showed significant lower accuracy than both other *device* conditions (Figure 8.6 a). Examining the interaction of *device* with *block* we found particularly strong learning effects for the mouse condition. The accuracy drawback of the mouse condition could efficiently be compensated through learning the task (Figure 8.6 b). A Tukey-test showed no further significant differences between devices in the last block of trials. In the second block only the difference between the Globefish and the mouse was found to be significant ($p < .05$).



(a) Sorted by *device* and *subtask*

(b) sorted by *device* and *block*

Figure 8.6: *Mean click error rate*

Our hypothesis on the superiority of position- over rate control could not be proved regarding the task completion times of the study. We suggest that the expected advantages could not be found, because the navigation task did not require much accuracy. Instead of targeted moves, participants rather scanned the environment continuously while trying to move fast. Following Anderson [7] this strategy is well supported by rate control. In fact, we observed that with the SpaceTraveller users tended to select moving targets with the mouse pointer while turning the viewpoint continuously. With position control instead motion input is necessarily interrupted by clutching operations to compensate the limitations of the physical input space. The Globefish with an isomorphous transfer function required even more clutching compared to the mouse. We assume that an accelerated transfer function could have improved its performance in our experimental task.

Estimated benefits of the mouse device could only be found in time efficiency. In terms of accuracy instead, we found advantages for two-handed interaction. Employing an additional input device for the non-dominant hand seems to enhance accuracy. The position-controlled Globefish however, showed larger advantages in that regard. We observed that with position control users worked more sequentially. Parallel input from both hands could rarely be recorded in the Globefish condition, but users altered between navigation input from the left hand and pointer input from the right hand. With the elastic rate controller instead, users tended to control viewpoint navigation and object selection in parallel. The results of our study indicate that sequential operation might decrease time efficiency, but improves accuracy.

8.6 Motion Parallax Study

To gather spatial orientation by looking around in a virtual scenery is an important task of many 3D applications but definitely not the only relevant one. Through interviewing and observing professionals working with CAD and DCC tools, we identified another important aspect of spatial navigation techniques: depth perception.

Most workplace set-ups for such applications do not provide a stereoscopic display. In wireframe mode, neither occlusions cues are available. Thus, to obtain depth information users need to rely on perspective as well as on motion parallax resulting from movements of the viewpoint and the manipulated objects. When working with unknown geometries and especially with organic shapes, perspective is not a very reliable cue. Motion parallax instead is very robust, but permanently requires viewpoint or object motion. Since usually no head tracking is provided, this can be achieved with input devices like those described in Section 8.4. Particularly challenging with respect to depth perception is the interaction with wireframe visualizations and control points that define the shape of complex 3D objects. Since the respective applications display wireframe lines and control points with a fixed size on the image plane, the user cannot rely on differences in size to distinguish their depth in 3D space.

8.6.1 Task

We implemented an evaluation scenario based on spatial control point selection. A number of yellow colored points were distributed on all six faces of a translucent cube, situated in front of the viewpoint. The cube could be translated and rotated in three dimensions without constraints using a 3D input device operated by the non-dominant hand. To complete the task, only small amounts of rotation were required, but we provided full 6-DOF interaction functionality to allow for different user strategies. The distance of the cube from the viewpoint had an important influence on the distance of the selection points in screen space. Keeping the object far away from the viewpoint resulted in short motion amplitudes for the mouse pointer between the points, but respectively harder differentiability between them. Moving the cube closer had the opposite effect. Reproducing the design of common software packages the square shaped points were visible and selectable from any point of view. They always maintained a fixed size of 2x2mm on the screen independent of their respective distance.

An arrow-shaped pointer, controlled by the mouse device in the dominant hand was used for point selection. Selected points switched color from yellow to red. One trial consisted of selecting four points defined by a surrounding square frame on one of the cube's six faces (Figure 8.7). The task was completed by having the four points correctly selected. Incorrectly selected points had to be deselected to complete the task. Hardly any perspective ever showed only the relevant selection points within the borders of the turquoise-colored frame. To recognize, which of the surrounded points really belonged to the relevant group, users needed to slightly turn the cube in most cases. Thus depth information to recognize the points located on the same surface as the framing rectangle could be obtained through motion parallax.

Each trial started with the appearance of the target frame, stretching over a quarter of one of the cube's faces. The distribution of selection points and the position of the target frame were randomized. Successful selection of the four points finished a trial and initiated the next one, starting with the appearance of a new target frame and the repositioning of all selection points.

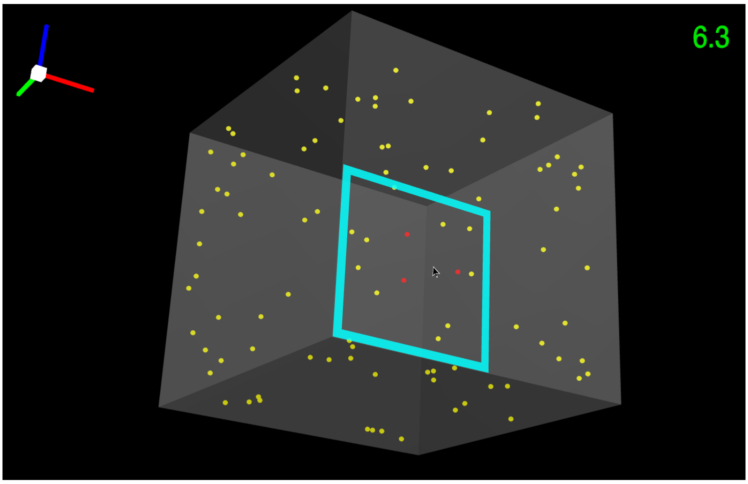


Figure 8.7: *A translucent cube covered with control points was presented to the participants of the study on depth perception. The four points framed by the turquoise colored frame had to be selected with the mouse pointer. Only through motion parallax the respective depth of these points could be recognized. Participants thus had to rotate the cube at least slightly using a 3D motion controller to the non-dominant hand.*

8.6.2 Apparatus

The experiments were conducted on a desktop set-up with the user seated in front of a 22 inch CRT display (resolution: 1920x1200 px), which showed monoscopic visual stimuli. The test application was running at a frame rate of 96Hz.

As in the study before, the mouse device assigned to the dominant hand might also be used for 3D manipulations if incorporating mode changes. The additional operation of full 3D rotation and translation with the 2D pointing device, however, requires at least four additional modes and results in the separation of inherently integral dimensions. Therefore, we compared only the 6-DOF devices for operation by the non-dominant hand.

For technical reasons we were required to use an earlier prototype of the Globefish device in that study. The employed mouse and SpaceTraveller devices as well as the respective transfer functions were the same as in the previous study, except that we decreased the control display gain for mouse input to seven for higher accuracy.

All elastic input with the SpaceTraveller device and translational input with the Globefish device were mapped to rate-controlled translations of the cube using a non-linear transfer function as in the study before. In contrast to the first study, position-controlled rotation inputs were now mapped to the virtual object using an accelerated transfer function. Previous experiences with the device indicated that an isomorphous mapping is less beneficial. We thus used a power function as described in [98] to enable precision as well as rapidity.

8.6.3 Hypothesis

We assumed that position control is the preferable choice to achieve spatial perception through motion parallax, even if controlled with manual input instead of head tracking. The required relative motion between object and viewpoint is minimal but the task requires high accuracy. Employing the Globefish to rotate objects of interest with position control rather than rate control as with the elastic SpaceTraveller should therefore result in higher time efficiency as well as higher interaction accuracy in the chosen task on spatial point selection.

8.6.4 Design and Procedure

To gain insights into expert performance on such tasks, we conducted three sessions on three consecutive days with each user and device, thus trying to ensure sufficient training on the task as well as on the tested devices. In each session, three blocks had to be completed with each device. The order of devices was balanced between two user groups.

Since the task was rather difficult for many participants, every test session included a training block for each device. Short breaks interrupted the blocks consisting of 16 trials. Overall, one session lasted about half an hour each day. After the third session a questionnaire was handed to the participants, asking them to report experiences during the tests and assess the tested devices on a five point Likert scale.

8.6.5 Participants

Twenty-two volunteers, aged between 20 and 33 years, participated in the study. All were students from different disciplines. Sixteen were male and six were female. All were right-handed. Sixteen participants reported to have experience with certain variations of the SpaceMouse device. Eighteen reported to be familiar with 3D computer games, seven reported experience with VR-systems and four among them were used to work with CAD or DCC tools. Four participants had no experience with 3D applications.

8.6.6 Results and Discussion

Data was entered into a 2 (device) x 3 (session) x 3 (block) analysis of variance (using Bonferroni adjustment for α in post-hoc tests) with order of devices as between-subjects factor. The order of the devices produced neither a main nor an interaction effect.

Regarding task completion times, the Globefish (4.75 s) significantly outperformed ($F_{1,20} = 15.66$, $p < .001$) the SpaceTraveller (5.04 s). From our observations we conclude, that the advantages of the Globefish stem from different interaction strategies with both devices. While rate control with the Spacetraveller encourages concurrent two-handed input, position control fosters sequential input which seems to enhance accuracy.

The Globefish also features a stronger distinction between translational and rotational input than the Spacetraveller. Accordingly we recorded twice as much simultaneous 6-DOF input in the Spacetraveller condition than in the Globefish condition, but this did not result in observable disadvantages. In either case only small displacements were applied to the cube and user's did not seem to have issues with keeping the cube inside the field of view.

A main effect on time efficiency was also found for *session* ($F_{2,40} = 8.47$, $p < .001$). Significant learning occurred between the first and the following two sessions ($p <$

.05), but post-hoc comparisons did not show a significant difference between the last two sessions (Figure 8.6.6). Learning effects became significant only between sessions (days) due to the training block performed before each session with both devices. No main or interaction effect on *block* could be obtained.

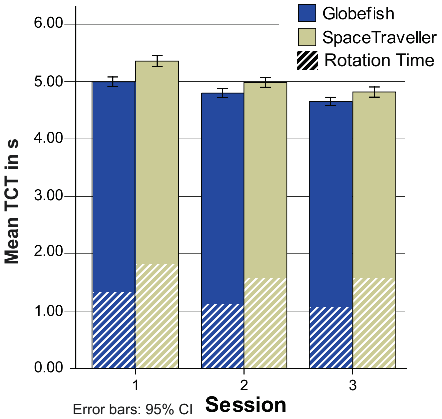


Figure 8.8: *Mean task completion times in seconds, sorted by device. The hatched partitions of the bars at the bottom illustrate the fraction of time, when the cube was rotated. The larger partitions correspond to the sum of idle times, object translations, and point selections without simultaneous object rotation.*

Additionally, we analyzed selection errors (wrong points selected) and click errors (missed targets) per trial. The marginal differences between device conditions on selection errors indicated that users were always able to discriminate the spatial location of points to select. Regarding click errors, that were counted when clicking outside a target, we only found a main effect of *device* ($F_{1,20} = 10.13$, $p < .01$). Rate-controlled rotations with the SpaceTraveller resulted in a significantly higher click error rate (24%) than using position control with the Globefish (20.75%).

As in the study before we observed that users tended to aim at selection points, while still turning the cube with rate-controlled input. This strategy of continuous object rotation, practically providing continuous depth cues through motion, resulted in the more complicated selection of moving targets. Sequential operation, as observed in the usage of the Globefish seems to be a better strategy.

Recorded operation sequences support this observation. We found that users spent about 41% more time on object rotations in case of the SpaceTraveller condition (figure 8.6.6 - Rotation Time). A much larger amount of time was dedicated to point selection than for turning the object in order to achieve depth perception. The more pronounced performance difference with respect to rotation only indicates a superiority of the Globefish device for that kind of task. Subjective ratings (from 1=best to

5=worst) confirmed a user preference for the Globefish (1.45) over the SpaceTraveller (1.73) for the spatial point selection task.

8.7 Conclusions and Future Work

We conducted two experiments on spatial navigation in desktop-based 3D applications. The results of the spatial overview task revealed two major results. Pure mouse-based interfaces provide comparable performance as using additional 3D navigation devices if only the most relevant degrees of freedom for viewpoint orientation need to be controlled. On the other hand we found significant benefits in accuracy through better balanced bi-manual input when using 3D input devices, which users of 3D modeling tools should take into consideration.

The experiment on depth perception indicated advantages in spatial perception for position-controlled over rate-controlled rotation input. The Globefish device performed significantly better than the SpaceTraveller with respect to time efficiency and accuracy. Although our first study suggests that for coarse spatial navigation rate control is well suited, we observed considerable advantages for the position-controlling Globefish in cases where accurate 3D object manipulations or exocentric viewpoint navigation were required.

The effects of input device characteristics on user performance remain a relevant research direction. It is the sensors and displays that constitute the tangible reality of virtual 3D environments. The relationship of the mechanical characteristics of an input device (shape, size, operation methods, etc.) to its usage in real-time graphics applications is a major aspect of that research. We suggest that the perception of the operation of physical devices provides users with important cues that can help to distinguish the states of interactive applications. For example, switching from position control to rate control should be always accompanied by swapping the mechanical device characteristics from isotonic to elastic. Future interaction devices could even allow for dynamic adaptations of their tangible characteristics to better suit the changing requirements.

Acknowledgments

We thank the participants of our studies as well as Susanne Günther and Sascha Gärtner for their tremendous support in implementing the evaluation tasks.

Chapter 9

Hold-and-Move – A Bimanual Cooperation Pattern

This chapter reports on joint work with Jan Dittrich and Bernd Fröhlich at Bauhaus-Universtität Weimar. It has been presented at ACM Mobile HCI 2012 and was published in the conference proceedings under the title:

"The Hold-and-Move Gesture for Multi-touch Interfaces"

Abstract

We present the two-finger gesture hold-and-move as an alternative to the disruptive long-tap which utilizes dwell times for switching from panning to object dragging mode in touch interfaces. We make use of a second finger for object selection and manipulation while workspace panning is operated with the first finger. Since both operations can be performed simultaneously, the cumbersome and hard-to-control autoscrolling function is no longer needed when dragging an object beyond the currently visible viewport. Single-finger panning and pinch zooming still work as expected. A user study revealed that hold-and-move enables faster object dragging than the conventional dwell-time approach and that it is preferred by most users.

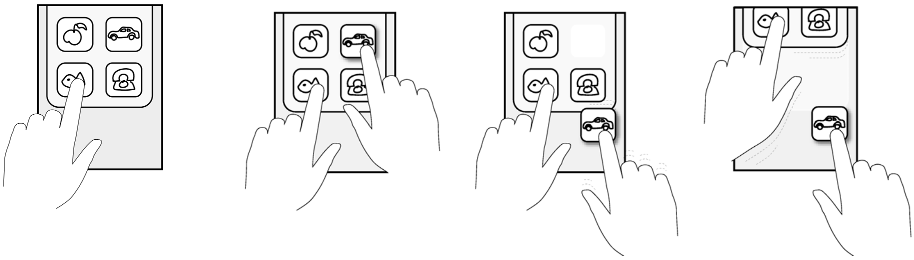

(a) Hold (b) Select (c) Move (d) Scroll

Figure 9.1: *The hold-and-move gesture. (a) The first contact point (here a finger of the left hand) is always associated with the background. (b) The second contact point (here a finger of the right hand) selects an item. (c) The item can now be moved in relation to the background that is held with the left hand. (d) The left hand may perform panning of the background and even clutching while the right hand holds on to the selected item.*

9.1 Introduction

Mode switching based on dwell times is widely used with mobile touch devices. It generally toggles between reference frames for motion input, e.g. panning or scrolling the entire interaction space vs. dragging individual items. Motion input from a single finger is primarily used for moving the entire screen content. Dragging individual items requires that they are first selected by a long tap based on a predefined dwell time. This can be very annoying. Dwell times often take too long if the user wants to switch to the respective mode and they may also occur accidentally.

We developed the hold-and-move interaction technique (Figure 9.1), which uses an implicit input differentiation based on Guiard's "left-hand precedence in action" principle [115]. Hold-and-move associates the first contact point with the background and thus motion input from a single finger always controls panning. In order to manipulate individual items, the background must be held with the first finger. A second finger may then select an individual item and move it in relation to the background. In addition, the first finger may perform panning of the background and clutching while the second finger holds on to the selected item. The manipulative gesture consistently ties the fingers to the graphics displayed under the contact point, which enforces the physicality of the interface.

The development of the hold-and-move pattern was motivated by an exploratory study on the timing of common multitouch input actions. We observed that dwell times between a touch event and the actual motion onset are an intrinsic characteristic of touch-based input and that their duration depends heavily on the activity. On average we recorded 130 ms dwell time during object dragging tasks if the functionality is directly available (without the requirement of an initial long tap). A regular long tap with a typical duration of 500 ms is therefore an interruption of the natural interaction flow. Simply reducing this threshold, however, is rarely feasible as it would increase the chance of false positives during navigation tasks.

The analysis of the pinch-zoom gesture confirmed the symmetric nature of this input action. Both fingers are placed down almost simultaneously with only 55 ms time difference on average. Hence, we designed the hold-and-move gesture as a complementary asymmetric two-finger gesture. It can be distinguished from symmetric input based on the time difference between both touch events. We found three main advantages of this approach:

- The disruptive long tap can be avoided.
- The asymmetric gesture is compatible with established one-finger gestures and can be clearly distinguished from symmetric two-finger input like pinch zooming.
- Manipulation and navigation can be operated simultaneously.

Mode switching based on dwell times provokes interferences between the two input modalities, e.g., object dragging and panning. Hold-and-move avoids this. The asymmetric two-finger gesture may instead interfere with other two-finger input, e.g., pinch zooming. Our studies indicate that this rarely happens as the pinch-zoom gesture is generally initiated symmetrically. A comparative user study revealed that hold-and-move can be more efficient than the dwell-time approach and that it is preferred by most users. We further analyzed the timing structure of the hold-and-move pattern and derive guidelines for its implementation.

9.2 Related Work

Guiard's analysis [115] on asymmetric bi-manual activity inspired the design of many highly effective human computer interfaces (e.g. [38, 48, 67, 167]). Unlike our approach, asymmetric bi-manual interaction techniques generally require the explicit identification of input from the user's dominant and non-dominant hand. Towards this end, several researchers recently proposed to operate tasks of the non-dominant hand with direct touch input while the dominant hand performs more accurate operations with a pen device [44,142,372].

Other approaches to increase the expressiveness of direct motion input build on explicit mode switching [205, 223], distinct hand shapes [81] or various motion gestures [249,291,292,369]. Li et al. [205] compared different mode switching techniques for pen-operated interfaces: dwell times, pressing the pen's barrel button, pressure and mode switching with additional input from the non-dominant hand. Dwell times showed the worst performance whereas the best results were achieved with bimanual input. This observation clearly supports our approach to circumvent dwell times by exploiting the additionally available information of multitouch devices.

Lank et al. [193] also analyzed mode switching with the non-dominant hand for pen input. They were focusing their research on the initiation pattern and found that the mode switching action triggered by the non-dominant hand generally precedes the dominant hand's action. Following the principles of naïve physics, Siio et al. suggested to use the palm of the pen-operating hand to fixate a virtual background for inking as if it were paper to write on [318].

Lü et al. [211] introduced a drawing-based selection technique for small targets on mobile touch devices. They note that drawing requires a mode change in most applications and suggest to hold the background with one finger before drawing with another one. While this idea is similar to our hold-and-move technique, it is also more constrained to serve only this particular application. The implementation details and the usability aspects of such a bi-manual mode switching technique are not addressed in their publication. In that sense, our work is complementary to theirs. We contribute an analysis of the timing parameters to distinguish hold-and-move from other two-point motion input, describe its general applicability and verify its usability benefits in a formal study.

9.3 The Timing of Touch Gestures

We were interested in the timing of common input actions on touch-screen devices. In particular, we analyzed unconscious dwell times in common panning and object-dragging tasks and also the time difference between both touch events initiating a pinch-zoom gesture. We expected that dwell times between a touch event and the respective motion onset would occur naturally, both in panning and dragging tasks. In case of occurrence, we were also measuring its duration to derive an optimal dwell time for mode switching. For the symmetric pinch-zoom actions requiring two instead of only a single contact point, we also expected intrinsic timing characteristics that would allow a clear distinction from other two point touch gestures.

9.3.1 Experimental Setup

We asked ten students from our campus to perform a series of dragging, panning and zooming tasks using a mobile touch device (Apple iPod®touch 2). All participants but one stated to have only marginal previous experience with touchscreen interfaces.

With detailed instructions and preparatory training we ensured that our test users were fluently operating the device during the data gathering. The purpose of the study was not explained to the participants.

Mobile devices are operated in different situations. On account of this we involved two *support* conditions in our experiments. All participants performed the tests while seated, half of the tests with the device held in their non-dominant hand (Figure 9.2 a, c, d) and the other half with only the dominant hand involved and the device fixed to the table surface (Figure 9.2 b).

For every touch input we recorded the time of the contact event and the corresponding motion onset. The latter was defined as a deviation from the contact position larger than 1.7 mm (11 px). In informal tests this position threshold was found to eliminate involuntary motion input while maintaining the responsiveness of the interface. Dwell times were measured as the time difference between these two events. For zooming gestures we computed the time difference between the contact events of both fingers. In the following we explain the experimental tasks and the involved variables in detail.

Dragging

Dragging an on-screen object always involves selecting it at its current position and moving it to the desired target position. We expected a short dwell time between these subsequent actions as both are aimed movements that require a certain amount of planning.

The *dragging* experiment consisted of a docking task that was designed in correspondence to the main menus of many touch-based mobile devices. A grid of gray squares was displayed with white outlines on a black background. One of these items was highlighted with a green fill color designating it as the object to be dragged. Another one or two squares with white fill color marked the docking targets (Figure 9.2 a).

We randomized the number of target options (one or two) and also varied the grid size (6x6 or 4x4 items) as such variations may affect the duration of dwell times. Each participant performed 120 docking tasks in four consecutive blocks, with short breaks of about one minute in between. If an item had been released outside the target area, it jumped back to its original position. The task had to be repeated in this case, which resulted in additional dragging actions for our records.

Panning

Depending on the users' intentions, panning a large map or scrolling a page may expose different timing structures. If a user is searching for a salient target or if the target distance is known, the motion behavior corresponds to aimed movement in pointing tasks [134]. On mobile devices this type of scrolling task is generally accomplished with a series of rapid strokes (flicks).

If users are scanning the displayed graphics or reading a text instead, panning operations are rather slow and continuous [7]. We expected that this difference between panning operations would also affect dwell times. Thus, we designed two panning tasks, both of which consisted of vertically scrolling a text on the display.

In the *Search* condition we instructed the users to scroll down until they recognized a text passage highlighted with a blue rectangle on a light gray background. The marker was set to 3 cm in width and 1.5 cm in height, thus easily recognizable even if the screen content was moving rapidly (Figure 9.2 b).

The *Read* condition on the other hand consisted of reading a short text of about 200 words (Figure 9.2 c). Thereafter some related questions were asked in order to motivate the users to read the article thoroughly instead of only skimming it.

Panning inertia as provided by the operating system was enabled during all tests. The line height was always set to 7 mm on the screen, such that 13 lines were be visible at once. The tasks were not repeated but they required a series of panning actions for completion. In the *search* condition, the target was 120 lines away from the initially visible workspace area, in the *read* condition the text stretched over 51 lines.

Zooming

Currently, zooming is the most relevant gesture on mobile devices that involves more than one contact point. Many other meaningful gestures can be performed with two fingers. However, the gestural vocabulary for mobile touch devices can only be extended with input patterns that can be clearly differentiated from the established ones. Toward a better understanding of the pinch-zoom gesture, we designed another experiment to capture and identify its timing characteristics.

The screen showed two rectangles, a smaller one containing a short text nested inside a larger one (Figure 9.2 d). The users were instructed to zoom in until the small rectangle fully filled the screen without cutting away any part of the contained letters. Then they were asked to zoom out until the minimal zoom level of 50% of the initial value was obtained and finally zoom back in to the initial state. This procedure had to be sequentially repeated for at least three times. Some participants did the task more often which provided us with a larger database for the evaluation.

(a) Dragging (b) Panning.Search (c) Panning.Read (d) Zooming

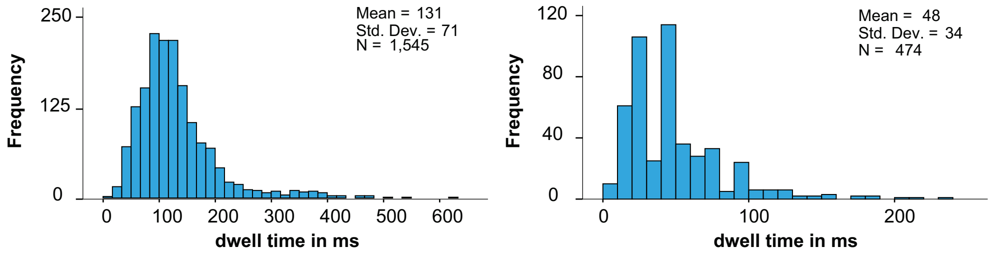
Figure 9.2: *Experimental tasks for measuring the timing of common touch gestures. The Search task (b) is shown in the support condition with the device fixed to the table surface. All others with the device held in the non-dominant hand.*

9.3.2 Recorded Time Intervals

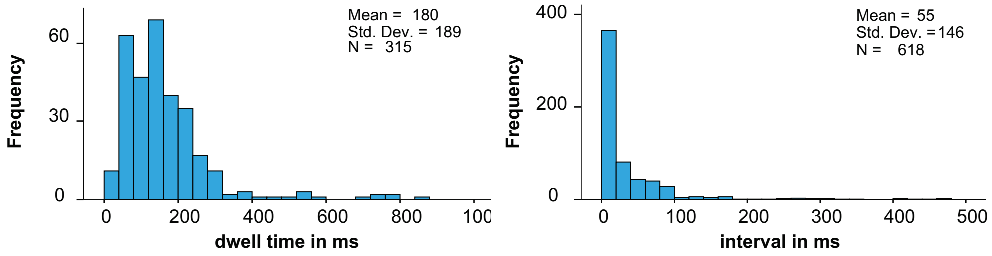
We recorded 1545 dragging actions, 315 panning actions in the *read* condition and 474 panning actions in the *search* condition. From this data we extracted dwell times between the contact event of the finger touching the screen surface and the corresponding motion onset. From the 618 recorded pinch-zoom gestures we derived the average time difference between the contact events of both involved fingers.

The results from the dragging and panning tests revealed that dwell times are indeed an implicit characteristic of touch input (Table 9.3.3). We also observed that their duration is clearly different among the tested tasks. The average dwell time during dragging tasks was 131 ms (*sd*=71 ms) (Figure 9.3 a). For panning tasks we measured average dwell times of 48 ms (*sd*=34 ms) if participants were scanning for salient targets (Figure 9.3 b), but if they were asked to read a displayed text with diligence, average dwell times increased to 180 ms with a standard deviation of 189 ms (Figure 9.3 c).

Regarding the pinch-zoom gesture, we observed that both fingers generally touch the surface almost simultaneously. We excluded 19 outliers from our records with a time difference of more than 1000 ms. In the other 599 cases the time differences were under 500 ms with a mean of 55 ms and a standard deviation of 145 ms (Figure 9.3 d).



(a) *Dragging* experiment (b) *Search* experiment



(c) *Read* experiment (d) *Zoom* experiment

Figure 9.3: *Histogram visualization of dwell times during the four experiments: (a) dragging, (b) search, and (c) read. Subfigure (d) shows the histogram of time differences between the contact of both fingers with the screen surface when initiating a pinch-zoom gesture during the zoom experiment*

9.3.3 Design Considerations

It seems quite obvious to exploit dwell times for implicit mode switching since they are an implicit characteristic of dragging tasks. However, we measured much shorter time intervals than the commonly used threshold of 500 ms. Dwell times above 500 ms occurred naturally in only 3 of all recorded dragging actions. In the other 1542 cases the dwell time would have been disruptive. Thus, reducing the threshold bears the potential to increase the fluency of interaction. During more than half of all recorded dragging actions a dwell time longer than 100 ms occurred. Hence, this value could be a better suited dwell-time threshold - if interferences with panning actions can be avoided.

Our measurements indicate that this may be possible in tasks that require only rapid flicks to navigate through the displayed content. Panning actions in the *search* condition exposed with 48 ms on average the shortest dwell times in our experiments. 95% of all rapid panning gestures we recorded in this condition involved a dwell time shorter than 110 ms (see Table 9.3.3).

percentiles	**min**	**5%**	**median**	**95%**	**max**
	dwell times				
dragging	11	46	111	270	624
search	4	12	42	110	237
read	10	44	141	452	2032
	interval between two contact events				
zooming	0	0	18	143	478

Table 9.1: *Percentiles of recorded time intervals during different input actions (all data in ms).*

On the other hand, the data recorded in the *read* condition shows that dwell times during panning actions significantly increase if the task is cognitively more demanding. While reading a text our test users induced dwell times of 180 ms on average. This is longer than the dwell times recorded during dragging actions. While reading, a threshold of 110 ms would have been exceeded in 63% of all recorded panning actions. In order to avoid involuntary mode switching for at least 95% of all the panning actions in the *read* condition, the threshold would have to be set to a value above 452 ms. This corresponds to the Apple®design guidelines for long press gestures. The default duration here is 500 ms. In our recorded data, this interval was exceeded in only 1% of all panning actions. We conclude that the common dwell time threshold of 500 ms is well adjusted to reduce the chance of involuntary activation. We also note, however, that it frequently interferes with the interaction flow during object manipulation.

In our study the pinch-zoom gesture exhibits a highly symmetric initiation pattern for most cases. The interval between the contact events of both fingers was less than 100 ms in 93% of the cases. The 19 excluded outliers show that pinch zoom is not necessarily initiated symmetrically, but in more than 98% of our records it was. We conclude that any input from two fingers that touch the screen with a time difference of more than 100 ms is most probably not meant to initiate a pinch-zoom gesture. It could thus be interpreted differently.

9.4 The Hold-and-Move Input Pattern

Li et al. [205] found that asymmetric bi-manual interaction can be an advantageous alternative to the disruptive long tap. Following the principles of Guiard [115], high-level motion input like navigation ought to be assigned to the non-dominant hand while the dominant hand performs object manipulation within the provided reference frame. Guiard also observed precedence of action of the non-dominant hand as a general principle of asymmetric bi-manual activities. We suggest that this principle can be directly exploited for the interpretation of multitouch input.

We implemented the hold-and-move gesture as follows: The first finger is always assigned to *hold* the background (Figure 9.1 a). A second touch input is used to select and *move* individual items relative to the background if the touch occurs at least 100ms after the first touch event (Figure 9.1 b & c). In the following, we refer to this time parameter as the hold time – as opposed to the dwell time to discern a long tap. If both touch events occur within less than 100ms their input is interpreted as a pinch-zoom gesture instead. An interaction sequence of hold and move only ends when both fingers release the touch sensitive surface. The assignment of fingers to background or foreground thus remains in effect if only one finger is released, which naturally enables clutching. Long-distance panning operations can be controlled with successive input of one finger while the other one keeps holding a dragged item. (Figure 9.1 d).

The mode switching technique clearly consists of two steps: holding the background followed by the selection of a foreground item. Hence, it should be easily distinguishable from the pinch-zoom gesture. Recall that we measured intervals shorter than 100 ms between both touch events in 93% of the 618 recorded pinch-zoom gestures. During this initial data gathering, only a single input mode had to be operated. We expected that the awareness of a second two-finger input mode would further increase the accuracy of the symmetric pinch-zoom initiation.

Target-based zooming is still available alongside with hold-and-move. Only the initiation of the input gesture matters for the differentiation between symmetric input like zooming (almost simultaneous contact of both fingers) and asymmetric input like hold-and-move (successive touch of both fingers). Thereafter, motion input may be induced symmetrically or asymmetrically.

We identify the following main advantages of using hold-and-move for mode switching:

- It eliminates involuntary object selection during panning as well as workflow interruption during intended object manipulation.
- The two-handed technique omits the need for automatic scrolling to drag an item beyond the visible screen area.
- It facilitates successive object manipulations. Once the background is attached to one finger, several objects can be sequentially dragged with another finger.

9.5 User Study

We evaluated the performance and user acceptance of both the long-tap and the hold-and-move mode switching techniques in two test applications using an Apple iPod®touch 4 device. Both test applications were focusing on object dragging. Only little workspace panning and no zooming was required for task completion. Nevertheless, both types of workspace navigation were continuously available during all tests in order to observe potential issues of interference: between hold-and-move and zooming as well as between object selection with the long-tap and panning.

Hold-and-move was implemented as described above with a 100 ms hold time threshold to discern the zooming functionality. The *long-tap* implementation used the common dwell time of 500 ms and provided autoscroll functionality that was actuated if the fingertip was less than 4 mm away from the screen border.

9.5.1 Experimental Tasks

Our two test scenarios (*List* and *Grid*), both consisted of several object manipulation tasks that also required some navigation input. The basic activity of moving icons or pictures across the screen was implemented as a color association task. The users were asked to drag colored items (red, green, blue, gray) to corresponding text fields naming the color.

If a dragged object was released on the incorrect target position, it was automatically moved back to its origin. After having moved a matching item to each visible target field, one set of tasks was accomplished and another version was loaded, proffering a different sorting of targets and pieces to be dragged.

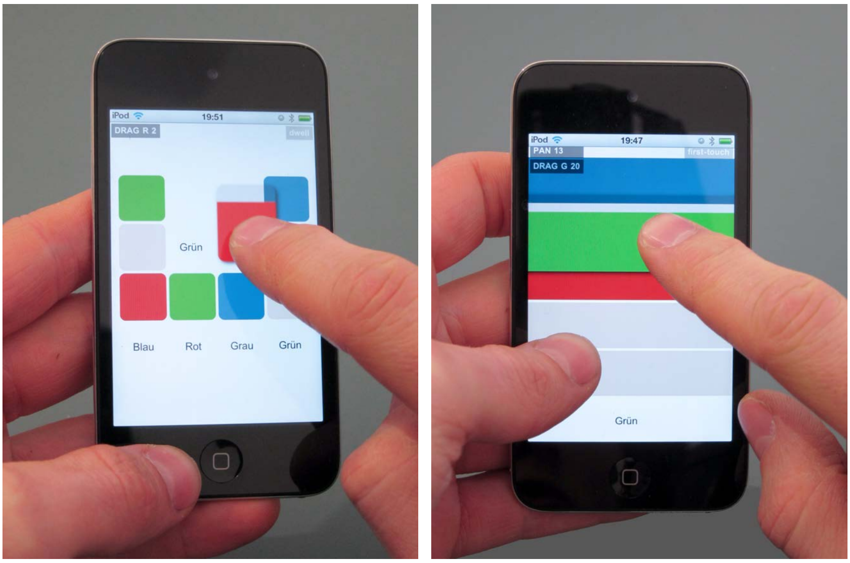
Note that the association from color to text added a constant cognitive load to the task. This was meant to reproduce a realistic situation in that users focus on accomplishing a certain task rather than the correct operation of the interface.

Grid

The color and text items were presented in a four by four grid structure (Figure 9.4 left). The first and third row each contained the four differently colored items. For each item there was a corresponding target field showing the name of the respective color in the line below. Besides this fixed vertical structure all items were placed randomly on the horizontal axis such that the users had to sort them while performing the task. Eight color items had to be dragged to the corresponding text fields on the line below. The subtask sequence was accomplished when all 8 items had been placed correctly. Then the next sequence was loaded.

List

The color items were presented in a list structure. The first three entries on top of the list always consisted of the color items red, green and blue in randomized order (Figure 9.4 right). Below, further entries contained gray items. All color items could be selected and moved across the screen, but only one red, green or blue item had to be dragged to the target position indicated by the word naming the respective color. We included seven different target distances (1–7 steps down in the list with Fitts's IDs ranging from 0.3–1), each appearing twice in random order. For the larger distances 5–7 the target position was sometimes too far away from the source to see both simultaneously on the screen. The users had to navigate to the target position in order to identify which item to drag there. Dragging the item to the target position correspondingly involved navigation. This second input mode was provided simultaneously in the *hold-and-move* condition or with autoscroll in the *long-tap* condition.



(a) The *Grid* task (b) The *List* task

Figure 9.4: *The test scenarios Grid ((a) with long-tap) and List ((b) with hold-and-move).*

9.5.2 Participants

20 people participated in the study, four of which were female, 16 male. The age ranged from 20 to 33 with a mean of 24.75. Six of the 20 participants claimed to use touchscreen devices on a daily basis, while 3 claimed to have no experiences with them. On a 5 point scale ranging from no experiences (1) to daily use (5) the mean was 3.77.

9.5.3 Design and Procedure

The participants were advised to hold the device in portrait orientation and in their non-dominant-hand. With each technique every participant performed six sets of eight dragging tasks in the *Grid* condition and thereafter 14 dragging tasks in the *List* condition. The order of techniques was counterbalanced. Each session started with a learning phase involving 6 dragging actions in a more holistic chess task that required also a lot of panning and zooming to manipulate the small chess pieces. No data was logged during the chess task, but it served the participants to subjectively explore the usability of the respective techniques. After all tests had been completed, we asked the participants to rate the suitability of both techniques for each scenario.

9.5.4 Hypotheses

We assumed that users would complete the dragging tasks faster in the *hold-and-move* condition – not because the dwell time in the *long-tap* condition is a disadvantage by itself but rather due to the users' difficulties to cope with the interruption of their workflow. In the *Grid* condition the dragging tasks were presented in sets of eight. This allowed users to take advantage of activity planning. In particular they could minimize the effort in the *hold-and-move* condition. Holding the background once enabled the object manipulation mode for all consecutive dragging actions. We thus expected a bigger advantage for *hold-and-move* in this scenario.

9.5.5 Results

Data was collapsed and entered into a 2 (*scenarios*)×2 (*techniques*) analysis of variance. Normality was verified using the Kolmogorov-Smyrnov test. We found significant main effects for task ($F_{(1,19)} = 56.0, p < 0.01$) and technique ($F_{(1,19)} = 62.7, p < 0.01$) as well as a significant interaction between these two independent variables ($F_{(1,19)} = 8.6, p < 0.01$). Post-hoc comparisons using Tuckey's HSD test revealed significant differences between the tested *techniques* in both *scenarios* ($p < 0.01$ for the *Grid* condition and $p < 0.05$ for the *List* condition).

As can be seen in table 9.2 the performance difference between both *techniques* is more pronounced in the *Grid* condition. A more detailed review of the data revealed that these additional benefits are indeed related to the improved integration of subsequent manipulation tasks. We observed that users were setting a new reference with an initiating touch event only twice within one set of the eight dragging tasks – just before starting to manipulate another group of four items in one line. Hence, the competitive edge for our novel gesture is less pronounced if we compare both techniques only in the first and the fifth dragging subtask: 2119 ms (*sd*=1267 ms) for *hold-and-move* vs. 2940 ms (*sd*=1565 ms) in the *long-tap* condition.

	Grid	**List**
long-tap	2866 (1459)	3459 (2151)
hold-and-move	1366 (991)	2651 (2395)

Table 9.2: *Mean task completion times in ms for both tested techniques (lines) in the two usage scenarios(columns). Standard deviation in parenthesis*

The impact of the target distance in the *List* condition was visible in the results, but no difference between *techniques* could be found in that regard. Overall we found considerable performance advantages of the *hold-and-move* technique as compared to the *long-tap* interface in both test scenarios (Figure 9.5).

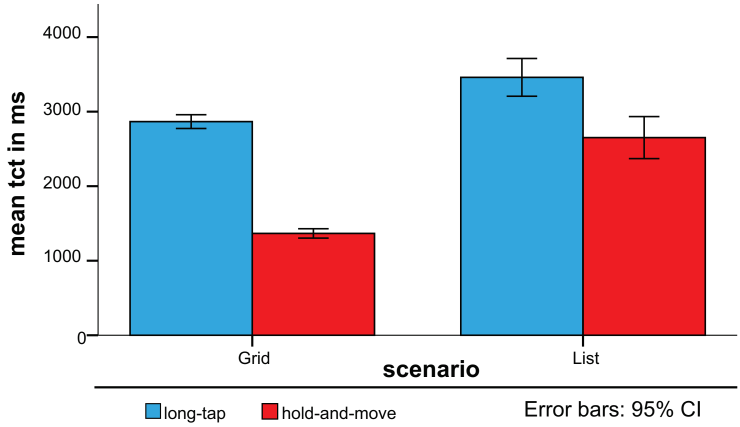


Figure 9.5: *Mean task completion times in ms*

Subjective user ratings reflect the quantitative results. The average usability rating of the two techniques on a 5-point Likert scale was similar with a mean of 4.0 for *hold-and-move* and a mean of 3.7 for the dwell times, indicating that both techniques were well suited for the given tasks. Most users had a clear preference for one of both techniques. Ten users preferred *hold-and-move* to operate the chess game while six voted for the *long-tap*. For the *List* sorting task eleven users preferred *hold-and-move* while seven users preferred the *long-tap*. Regarding the *Grid* task, many users appreciated the benefits of *hold-and-move* for successive object manipulation. Sixteen of 20 users preferred the novel technique in the *Grid* condition while three voted for the *long-tap*. Only few users were undecided in some scenarios.

9.5.6 Error Analysis

Different types of erroneous input may occur with the two mode-switching techniques at test. Dwell-time-based object manipulation may interfere with other motion input from one finger, e.g. panning. Hold-and-move, on the other hand, can interfere with zooming that is also operated with motion input from two fingers. In both cases, the correct interpretation of the user's input critically depends on the timing.

Quantifying such mode switching errors is not trivial. Without a predefined input sequence, we cannot easily know which mode changes were intended and which were not. Additionally, we must consider false positives (the mode changed although not intended) and false negatives (the intended mode change was not achieved).

False Positives

From our initial study on the timing of touch gestures we derived that a dwell time of 500 ms is a well suited threshold to minimize false positives during panning. Only 14 of the 789 recorded panning actions involved a dwell time longer than 500 ms.

We also estimated the rate of false positives for the hold-and-move gesture from the records of pinch-zoom gestures gathered during this initial study. In 7% of the 618 recorded pinch zoom gestures, both fingers touched the screen surface with a time difference longer than our threshold of 100 ms. They would have had invoked the object manipulation mode although their input was meant to control the zoom level. We expected, though, that users would initiate the pinch-zoom gesture more accurately with both fingers simultaneously if they are aware of a second two-finger gesture that is initiated asymmetrically instead.

During our second study comparing hold-and-move to dwell-time based mode switching we recorded 483 hold-and-move gestures, 446 of them were completed with the successful docking of one or more dragging items. In the other 37 cases the gesture may have been initiated involuntarily. This is plausible if both touch events occurred within the range of time intervals we measured for pinch-zoom gestures.

We define this range based on the average interval we measured for pinch-zoom gestures plus three times standard deviation (55 ms + 3 ∗ 145 ms = 493 ms). In 18 cases of our 37 error candidates both fingers touched the screen within a shorter interval, which would correspond to a rate of 3.7% false positives for a hold-time threshold of 100 ms.

False Negatives

Towards an estimate of false negatives, we can compare the frequency of "unnecessary" input actions. Most of our experimental tasks did not require panning or zooming as the source and the target position were displayed on visible screen area. Any panning or zooming would thus be inefficient and can be interpreted as an error.

In the *Grid* task users could benefit from the integration of up to eight subsequent dragging operations in one hold-and-move sequence while the long-tap involves a mode selection for each subsequent dragging operation. Thus, we only consider the first dragging action from each of these sets for the error analysis. Some of the *List* tasks involved target positions off screen and therefore required panning or zooming. This should have the same impact on the performance with both techniques, but to account for this difference we distinguish short and long target distances in the analysis. Distances of up to 4 rows were always defined to be short since the targets were always visible without navigation.

Zooming occurred rarely in all tasks and with both techniques (see Table 9.3). The slightly higher number of zooming actions in the *hold-and-move* condition, indicates that involuntary input may have occurred, but none of the small differences proved to be statistically significant. We conclude that hold-and-move did not had a relevant impact on the operability of pinch zoom during our experiments. Although we applied a rather low threshold of 100ms, only one of 20 users noticed the possibility of interfering with pinch-zoom input.

For mode selection based on dwell times false negatives are indicated by an increased amount of panning actions. We observed that the average number of panning actions per docking task is more than two times higher in the *long-tap* condition as compared to *hold-and-move* (see table 9.3). A MANOVA reveals that the rate of panning actions differs significantly between techniques ($F_{(1,19)} = 25.68, p < 0.01$) and also between the three groups of tasks ($F_{(2,38)} = 30.06, p < 0.01$), while the interaction between both factors is not significant. We conclude that indications of input errors related to dwell times can be found in all task conditions. User feedback supports this observation. Five participants explicitly complained about being interrupted by dwell times.

	Grid	short List	long List
	zooming input		
long-tap	0.13 (0.47)	0.12 (0.49)	0.26 (0.64)
hold-and-move	0.13 (0.59)	0.07 (0.28)	0.19 (0.61)
	panning input		
long-tap	1.62 (1.01)	1.66 (0.95)	2.98 (1.405)
hold-and-move	0.58 (1.59)	0.42 (0.97)	2.08 (3.028)

Table 9.3: *Average numbers of panning and zooming actions per dragging task. Standard deviation in parentheses.*

9.5.7 Hold-Time Analysis

With a minimal hold-time threshold of 100 ms, false negatives do not seem to occur at all. Instead, we observe a chance to invoke the object manipulation mode involuntarily. This risk can be reduced by maximizing the applied threshold. Our records from 446 successful hold-and-move gestures, indicate that this is a feasible approach. The average hold time was 658 ms (*sd*=485 ms). Interestingly, this is longer than commonly applied dwell times of about 500 ms, which confirms our assumption that the reason for the inferior performance of the *long-tap* is not the dwell time itself. User feedback indicates that the drawback rather results from the difficulties to cope with the workflow interruption.

Only 5% of the hold-and-move gestures were executed with a hold time of less than 190ms (Table 9.4). We reasoned that such extremely short hold times only occurred because our study was focused on consecutive dragging tasks. Consequently, we suggest a hold-time threshold above 150ms (corresponding to the 95% percentile value of pinch-zoom input) and below 190ms (corresponding to the 5 percentile value of hold-and-move gestures). Depending on the target user group and different applications, other values may offer a better suited balance (see Figure 9.6).

percentile	**min.**	**5%**	**median**	**95%**	**max.**
hold time	103	189	514	1681	2927

Table 9.4: *Hold time extrema (all data in ms).*

9.5.8 Discussion

The results of our user study demonstrate that the hold-and-move gesture is an efficient alternative to the disruptive long-tap. Note that we tested with a very basic implementation of the technique. In practice, we expect that applications will take further information into account. Most obviously, context information should be included as known from using the long tap for object selection, which is only executed if the finger rests on top of a selectable item. In applications with a sparse distribution of selectable items, this simple adaptation would directly enable to operate zooming even after asymmetric initiation.

Compared to the long-tap, hold-and-move has the limitation in that it cannot be operated using only the thumb when holding the device in one hand. This will be less of a restriction for larger devices that are generally not operated single-handedly, like tablet devices. In addition, there is no technical reason that one-finger dragging based on dwell time cannot be made available in the interface in addition to the two-finger hold-and-move gesture.

It is important to consider that the normal mode of interaction for both techniques is navigation via panning and scrolling. We tested them both in manipulation-focused tasks. We expect that hold-and-move is beneficial also for navigation tasks that do not require a secondary input modality, since it prevents accidental mode switching (false-positives). Hold-and-move demands a second touching finger and cannot be triggered involuntarily in single-touch navigation (e.g. if leaving the finger on the screen while reading).

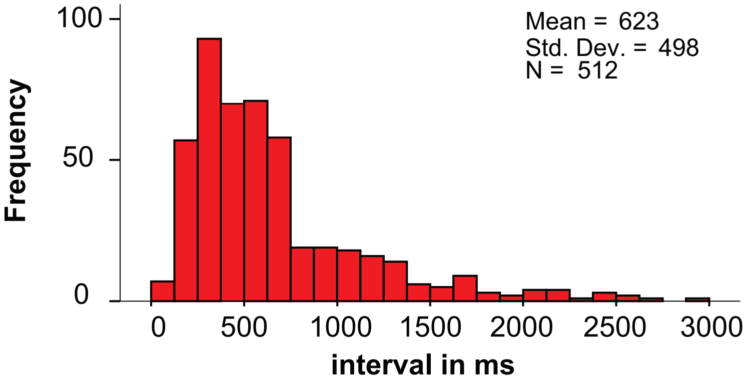


Figure 9.6: *Histogram of recorded time differences between the contact of both fingers with the screen when initiating a hold-and-move input gesture.*

9.6 Conclusions and Future Work

We designed the two finger object dragging pattern hold-and-move to avoid the disruptive long-tap in touch interface implementations. Our approach involves a second finger to enable the smooth execution of object selection followed by object dragging. With the novel technique regular workspace panning remains available during dragging. As a side effect, the cumbersome and hard-to-control workspace autoscrolling function is no longer needed when dragging an object beyond the currently visible viewport. Involuntary object selection during panning is also eliminated. Our user study revealed that the hold-and-move pattern allows for faster object dragging than the conventional dwell-time approach and is preferred by most users. We also showed that the novel gesture can be applied without affecting the recognition quality of the established single-finger panning and pinch-zoom gestures.

Pinch zooming mimics squeezing and stretching the screen content. Consequently, the finger motion is induced along the vector between both initial touch positions. Instead, for hold-and-move motion input is induced in the direction of a target position that is independent of the orientation of the initial touch positions. This observation could be used to further reduce involuntary mode changes. The implementation of the hold-and-move pattern could also benefit from the possibility to adapt the thresholds to the user and from context sensitivity. For example, if there are no objects near the second finger that can be dragged, the user is likely to start a pinch zoom even though the threshold for enabling this gesture has been exceeded.

While object dragging may be only rarely required in common mobile phone applications, the mode-switching technique enables many other functionalities that are similarly operated. Selecting snippets of text and images from websites and pasting them into a new document is currently a cumbersome task. Hold-and-move bears the potential to operate such tasks with comparable productivity as with the mouse in desktop environments. Furthermore, it provides a reliable basis for drawing-based interaction techniques like Gesture Avatar [211].

Our investigations into the timing structure of common multitouch gestures and of our hold-and-move pattern provide a solid basis for designing and implementing further multitouch gestures that do not require any additional hardware. Such gestures can be easily integrated into multitouch enabled operating systems for mobile devices, which would extend the slowly evolving multitouch language.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF) under grant 03IP704 (project Intelligentes Lernen).

Many thanks go to Felicitas Höbelt and Daniel Fischer who provided the software for the test applications. We also thank the participants of our studies and the reviewers of this paper for their helpful comments.

Chapter 10

Multi-User Virtual Reality

This chapter reports on joint work with André Kunert, Stephan Beck, Roman Reichel, Roland Blach, Armin Zink, and Bernd Fröhlich at Bauhaus-Universtität Weimar. It has been presented at ACM Siggraph Asia 2011 and was published in the conference proceedings under the title:

"C1x6: A Stereoscopic Six-User Display for Co-located Collaboration in Shared Virtual Environments"

Abstract

Stereoscopic multi-user systems provide multiple users with individual views of a virtual environment. We developed a new projection-based stereoscopic display for six users, which employs six customized DLP projectors for fast time-sequential image display in combination with polarization. Our intelligent high-speed shutter glasses can be programmed from the application to adapt to the situation. For instance, it does this by staying open if users do not look at the projection screen or switch to a VIP high brightness mode if less than six users use the system. Each user is tracked and can move freely in front of the display while perceiving perspectively correct views of the virtual environment.

Navigating a group of six users through a virtual world leads to situations in which the group will not fit through spatial constrictions. Our augmented group navigation techniques ameliorate this situation by fading out obstacles or by slightly redirecting

individual users along a collision-free path. While redirection goes mostly unnoticed, both techniques temporarily give up the notion of a consistent shared space. Our user study confirms that users generally prefer this trade-off over naïve approaches.

10.1 Introduction

3D television sets and 3D cinemas display only a single stereoscopic image stream, and thus there is only a single location from which a person observes a perspectively correct view of the displayed scenes. All of the other spectators perceive the 3D scene more or less distorted, which inhibits precise spatial perception of the displayed geometry. While this may not matter much in movie theaters, these distortions significantly hamper the acceptance of 3D technology in many other application areas. In order to compensate for this, each user must be provided with individual stereoscopic image pairs, which are rendered for the exact position of the user in front of a display. While the computing power to generate multiple views of interactive content is available, the display technology for presenting large individual stereoscopic images for multiple users is still lacking

We developed the C1x6, a new projection-based stereoscopic display for six users (Figure 10.1). Our system consists of six customized DLP projectors, each of which projects six fast time-sequential images in one of the primary colors. By differently polarizing the light output of the first set of three single color projectors (red, green, blue) than those of the second set, we are able to project twelve separable full-color images onto a projection screen. Our intelligent high-speed shutter glasses can be fully controlled from the application level. This feature is used to keep the glasses open if users look away from the screen or for supporting a VIP high brightness mode if less than six users are involved. We developed the software and hardware infrastructure to generate, warp and feed the stereoscopic images for the six tracked users into the projectors.

Multi-user displays enable co-located collaborative work in shared virtual environments. For collaborative design reviews we developed the Spheron, an input device which makes interactions transparent to other co-located users. However, when navigating a group of users through a virtual building, many situations arise in which there is not enough space to place the users in the virtual world in the same way as they are positioned relative to each other in the real world. This problem did not exist in common projection-based virtual reality systems, where all the observers share the same perspective as the head-tracked navigator. Therefore, we present several approaches to facilitate group navigation in such situations by avoiding collisions of group members with surrounding objects such as walls and other obstacles.

The main contributions of our work fall into three areas:

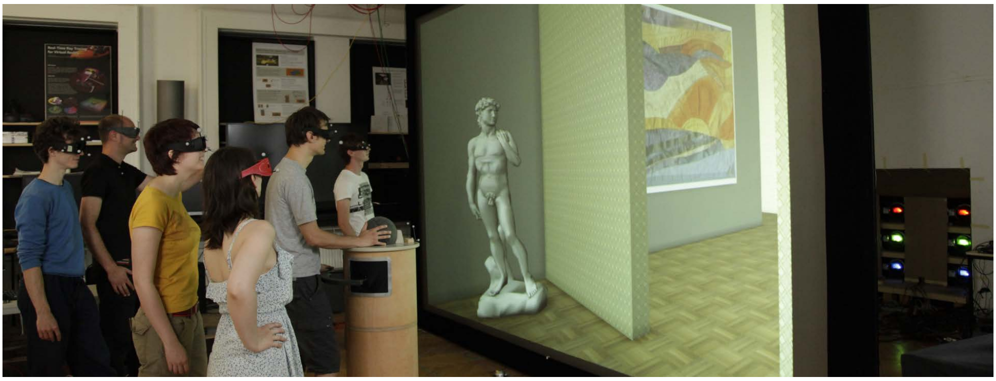


Figure 10.1: The six-user projection system. The lenses of the six projectors appear as bright color spots on the right side. Each of the six users is tracked and provided with a perspectively correct image. The Spheron, our group navigation device, is centrally placed in front of the display. 12 different images are projected onto the screen, only one image is shown here.

- A synchronized DLP projector array that is capable of displaying twelve high resolution (1920x1200) full color image streams at 360Hz—60Hz per user. Left and right eye images are simultaneously projected and separated by polarization.
- Application-level programmable shutter glasses consisting of double-cell liquid crystals that enable intelligent shutter control and provide fast switching speeds as well as high-contrast.
- Augmented group navigation techniques that avoid collisions when traveling through narrow spaces. Our user study reveals that these techniques are preferred over naïve approaches.

Besides these central contributions, significant amounts of engineering are necessary to build and run such a complex system, including a custom digital video-multiplexing hardware, synchronization of all the components and real-time color convergence through image warping. Measurements show that our six-user system achieves almost the same brightness per user as a stereoscopic single-user display based on the same type of projector would.

10.2 Related Work

The most straightforward way to provide multiple users with individual views of a shared virtual world is the use of personal displays such as head-mounted displays

(HMDs) or handheld displays. The Studierstube system by Schmalstieg et al. [300] supported collocated collaborative augmented reality using see-through HMDs. [12] performed a collocated architectural design review by providing multiple users with HMDs. [150] equipped multiple users with head-mounted projectors, which projected onto retro-reflective walls.

Projection-based stereoscopic displays such as the CAVE [65] have a long tradition in the virtual reality domain, but there have only been a few approaches providing multiple tracked users with individual stereoscopic images. The two-user Responsive Workbench [4] displays four different images in sequence on a CRT projector at 144Hz, which results in 36Hz per eye per user. They also developed custom shutter glasses for cycling between four eyes. This system was the first demonstration of a two-user system, but suffered from flicker, low brightness and crosstalk. Barco developed the "Virtual Surgery Table", which provides two users with stereoscopic images by differently polarizing the light output of two active stereo projectors [24]. This approach was also used for a large projection wall described by Riege et al. [282]. All these systems are limited to two users and cannot be easily extended to support more users. Bolas et al. presented a modified DLP projector, which is capable of running at 120Hz [40]. They also briefly mention the integration of such single-chip displays with the optics of a 3-chip DLP system to achieve a three-user system running at 60 Hz per eye in a single projector. Unfortunately, no technical details were provided. Our approach is similar in that it also uses customized 120Hz projectors, but we realized a complete fully synchronized six-projector system and demonstrate its use for six users.

Kunz et al. employed a pair of shuttered LCD projectors to generate an active stereo display for their blue-c system [112, 191]. Fröhlich et al. extended this approach to support four users by using eight shuttered LCD projectors [99]. To limit the shutter frequency, they used shuttering to cycle among the users and polarization for separating the left and right eye images. However, shuttering projectors is not a very light-efficient approach considering that each projector is blocked for most of the time. For example, in the four user setup, each projector is blocked for three-quarters of the time and thus 75% of the light output is lost. This is also the reason why it does not scale well to more users.

There are also a number of special purpose multi-viewer displays. The PIT [10], the Illusionhole [174] and the Virtual Showcase [39] use different approaches to assign a separate partition of a projection screen to each user, where the stereoscopic images for each person are displayed. The PIT uses two orthogonal screens, in which each user looks at only one of the screens. The Illusionhole uses a circular mask on top of a tabletop projection. By looking through the mask the users positioned around the table see different areas of the screen, where their individual images are presented. The Virtual Showcase consists of a tabletop projection with a truncated half-silver mirror pyramid (or cone) placed in the middle of the table. By looking into the mirror, users positioned around the table see a reflected image off the tabletop. The stereoscopic images are rendered such that the virtual objects appear inside this Virtual Showcase. The Joint Space Station [240] uses a similar approach based on separate Virtual Workbench displays (cf. [272]) facing each other. These displays are limited to two to four users with a small overlap of the users' viewing frustra, which considerably limits the size of the objects that can be displayed in the shared virtual space. Maksakov €t al. simply used separate viewports on a larger screen for each user to provide individual head-tracked monoscopic views of a 3D scene [219]. Such approaches effectively discard the notion of a locally shared space and require similar interaction techniques and affordances as do distributed multi-viewer systems.

Dodgson [76] and Favalora [83] provided an introduction and overview of the many other types of multi-view displays, in particular autostereoscopic and holographic systems. While the use of such technology for displaying large, interactive and full color 3D images for multiple non-stationary users is the ultimate goal, all of these systems pose different limitations. However, various recent developments are convincing solutions for particular application domains. Cossairt et al. [63] and Jones et al. [165] developed similar approaches for occlusion-capable parallax multi-view 3D displays. Their systems use modified DLP projectors to project fast time-sequential images onto a rotating anisotropic projection surface. Both systems achieve about one degree of angular resolution and support a 180 and 360 degree field of view, respectively. Jones reported on a further refined prototype of such a system and showed its use in a very convincing real-time one-to-many teleconferencing application [163, 164]. Due to the use of a single projector, the bandwidth of these systems is limited, which results in a small color depth of one bit color or even only black and white depending on the used DLP projector type. In addition, such a system is difficult to scale to a larger size due to the rotating display surface.

The research surrounding collaborative virtual environments (CVEs) has mostly focused on distributed collaboration (see, e.g., Benford et al.'s review of the history of CVEs [28]). Otto et al. [252] and Wolff et al. [368] provided a solid analysis of the requirements for supporting closely coupled collaborative tasks in a shared virtual workspace for non-collocated users, which also apply to a certain extent to collocated collaboration. However, there is limited work on collocated collaboration in projection-based multi-user virtual reality. The original two-user Responsive Workbench work [4] suggested the use of specialized views, which were used to provide different information to each user, as in a teacher-student scenario. Riege et al. suggested the use of a bent pickray to visualize the constraints that are involved when two users are jointly manipulating an object with six degrees of freedom [282]. D'Angelo et al. showed that stereoscopic display in combination with collaborative manipulation improve task performance and are clearly preferred in a complex assembly task involving two users [72]. Argelaguet et al. demonstrated the use of specialized views to reduce the problem of interpersonal occlusion [8]. All these approaches consider only two collaborating users and focus on joint manipulation. It is not clear how these approaches scale to more users.

Bowman et al. provide an overview and introduce a taxonomy for the large variety of navigation techniques for virtual environments [43]. However, the problem of navigating multiple collocated users with individual views through a shared virtual world has not yet been addressed. Group navigation as it is defined here – moving multiple people simultaneously through a virtual environment – is a new problem that is closely linked to the introduction of stereoscopic multi-viewer systems. Augmented group navigation techniques to mitigate associated issues are orthogonal to general single-user navigation techniques. In our setup each head-tracked person can independently walk in front of the display, but apart from that, does not independently travel within the environment since otherwise the group would no longer share a consistent virtual space.

10.3 Synchronized 12-View Projector Array

Our goal was to build a fast time-sequential full color DLP-based system which also exploits polarization. Our approach is based on the following ideas:

- Color wheel DLP projectors project the different primary colors as fast time-sequential images. There are various color wheel versions; we assume a basic three-segment color wheel consisting of three color filters, one in each primary color: red, green and blue. If the color wheel is removed, we can project three monochrome time-sequential views (Figure 10.2) instead of the different primary colors of a single view. By using three projectors and equipping each projector with a primary color filter, we regain full color images for three views.

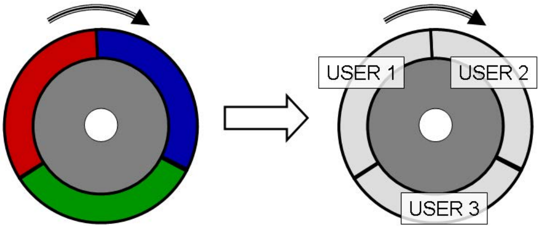


Figure 10.2: *A three-segment color wheel. We display individual images for three eyes instead of time-sequential colors.*

- Most DLP projectors rotate the color wheel at least twice per video frame and are thus effectively running at 120Hz while 60Hz input is provided. However, at the time of our development, a 1920x1200 pixels resolution projector was not yet available, which would accept a 120Hz stereo signal. Thus we had to extend an existing projector to process a 120Hz image stream or to interleave two 60Hz streams. This way we could project six different views at 360Hz (three views times two rotations times 60Hz).

- Polarization can be effectively used in combination with shuttering to double the number of views, thus allowing 12 views to be achieved using two times three projectors.

Such a system maintains the brightness of a single user active stereo system since we are using six projectors for six users. In addition, we retain full color depth, full resolution (1920x1200) and a 60Hz refresh rate. Figure 10.3 shows an overview of our setup.

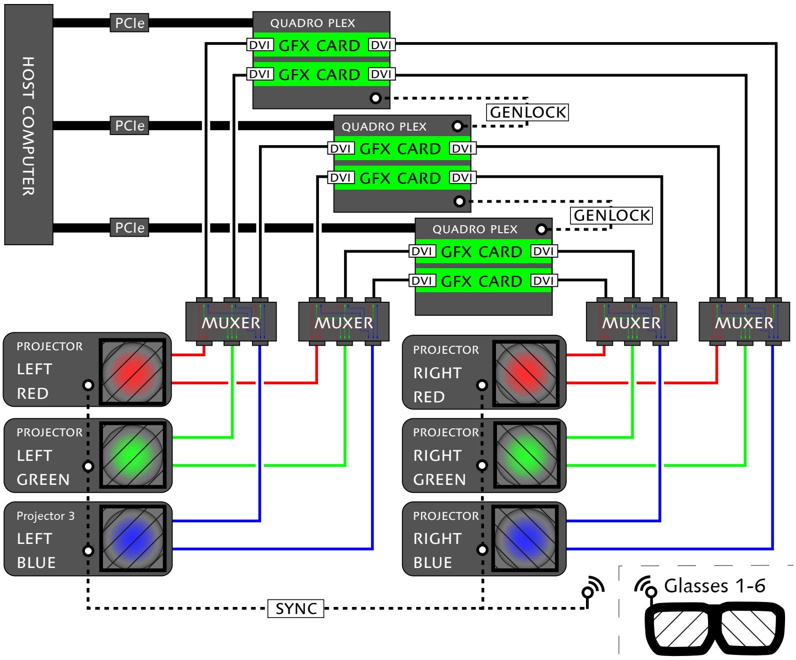


Figure 10.3: *The projector array is driven by a single computer. Three synchronized NVIDIA Quadro Plex 7000 graphics systems are connected to the host computer via separate PCIe interfaces. Each Quadro Plex consists of two graphics cards with two DVI outputs each. It produces the left and right eye images for two users. Sets of three DVI outputs carrying the images for three eyes are connected to the video multiplexers (muxer), which rebin the image streams by color and send them to the respective projectors. The left three projectors display the left eye images for the six users, while the right three projectors display the right eye images. The two sets of projectors emit differently polarized light which matches the polarization of the users' left and right eye shutters. External synchronization is provided to the projectors and to the radio-controlled shutter glasses.*

In building our 12 view projection system, we had to develop the following main components:

- A 360Hz projector with external synchronization capable of projecting six different views.
- Video-multiplexing hardware for feeding each projector with six different views, which are generated on different graphics cards.
- Real-time image warping for geometric alignment of images, since the color components of the images are projected from different projectors.
- Custom radio-controlled shutter glasses that provide fast switching an high contrast.

10.3.1 The 360Hz Projector

We modified six *Projectiondesign F32* DLP projectors to accept two 60Hz input streams and alternate the display of these image streams at 120Hz using an approach similar to the one presented in [147]. The F32 projector contains a separate input and scaler board, which accepts the DVI image stream and scales it to the resolution of the Digital Micromirror Device (DMD). The resulting 1920x1200 images are sent on to the formatter board, which reformats the image stream and sends it to the DMD. The main modification was the addition of a second input and scaler board, as well as a formatter board. The second input pipeline accepts a second image stream. From the formatter boards, the image stream is sent to the DMD via an LVDS interface. We added an LVDS switch in front of the DMD, which is connected to the two formatter boards for switching between the two image streams as shown in Figure 10.4. In our experiments we found that the F32 projector uses exactly the same DMD patterns for the first and second rotations of the color wheel. Thus there is no reduction in color depth involved by using two different input streams. Our projectors have the basic three-segment color-wheel firmware installed.

Doubling the input and scaler boards and switching between these two input streams is necessary if the projector accepts only a monoscopic image stream at e.g. 60Hz. This major hardware modification is not needed if a stereoscopic projector is available, which uses alternating rotations of the color wheel for projecting the left and right eye images. However, it also needs to have a three-segment RGB color wheel firmware installed for mapping each of the primary colors to a different user.

A dedicated IO board is used in the F32 projector for taking care of the fans, lamp power control and temperature sensors. IO status is routed from the IO board through the formatter board to the input and scaler board. We developed a microcontroller board for interfacing with the IO board, and doubling its inputs and outputs to keep the second input pipeline alive. The microcontroller operates the crossbar and its power supply.

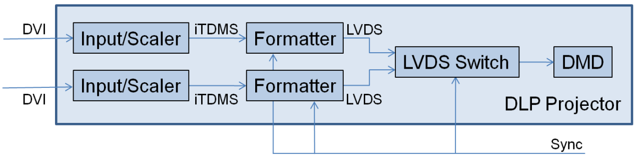


Figure 10.4: *Main electronics components and signal flow inside our modified DLP projector. Two synchronized 60Hz DVI signals are fed into the projector. The input and scaler board adjusts the size and refresh rate of an input stream to the requirements of the formatter board and sends it via an iTDMS connection to the formatter board. The formatter board formats the bitstream for the DMD chip and sends it to the DMD via an LVDS link. The LVDS switch alternates between the two LVDS streams from the formatter boards. The sync signal is externally generated and keeps all the components in sync.*

The last major modification of the projector involves the color wheel. Since the F32 projector uses two lamps, there were also two color wheels. Since we only project a single color per projector, we removed the color wheels and use a fixed color filter in one of the primary colors. Cooling and infrared filtering had to be added to avoid heat problems. Color wheel projectors receive their internal sync from the rotating color wheel. We provide an external synchronization signal to all the projectors to keep them in perfect sync. The same sync is also provided to our custom shutter glasses.

10.3.2 Video-Multiplexing and Geometric Alignment

We built a digital video-multiplexer hardware, which takes three full color image streams from the graphics cards, rebins them by color and sends them to the respective projectors. We use four of our video-multiplexer units to route the color components of the 12 DVI streams to the respective projectors (Figure 10.3).

Since the primary colors are displayed by different projectors, the colors of the projected images will be misaligned. Although shift lenses and software keystone correction can help in some configurations, the overall image quality is not acceptable for our needs. Therefore we use a real-time image warping technique that warps each pixel of a rendered image to the correct location on the screen. The multiplexers route each image to three different projectors and thus each color component has to be warped to a different location to achieve color convergence. This warping step is based on individually precomputed look-up tables, one for each projector, and is performed in real-time as a post processing step in a fragment shader program.

Our algorithm automatically calibrates multiple projectors onto a single target projector. This involves several computation steps. First, we capture Gray code patterns with a monochromatic 5 megapixel camera to derive camera-to-projector maps for each projector, similar to [69]. Based on these maps, our algorithm computes the largest projection area that can be covered by each projector. Finally, look-up tables are computed for each projector so that every location within the projection area maps to a location of a rendered image. In this calibration process, the acquisition of the Gray code patterns is the most crucial step. For achieving a good signal-to-noise ratio, our calibration tool simultaneously renders the Gray code patterns on all six graphics cards contributing to the projection of a single projector. We achieve a precise per-pixel alignment of the six projector images throughout the projection area.

10.4 Intelligent Shutter Glasses

Shutter glasses are unavoidable in our approach. They need to work at 360Hz and the left and right eyes need to be differently polarized. Since shutter glasses consist of two crossed polarizers with liquid crystal material in between, the left and right eye shutters just need to be rotated against each other by 90 degrees to achieve orthogonal filtering capabilities. Regular liquid crystal shutters are not suitable for our system since they have asymmetric opening and closing properties. They close quickly (e.g. less than 0.2ms depending on the operating voltage) and open slowly (e.g. longer than 2ms), too slow for 360Hz cycles. The standard solution to circumvent this problem is the use of ferro-electric (FLC) shutters with symmetric opening and closing times of less than 0.1ms. While such a solution is easily capable of running at 360Hz, FLCs are much more expensive than standard liquid crystal (LC) shutters and they are very fragile. FLCs are also designed to work with symmetric open/close timings, which is not the case in our setup.

As an alternative to FLCs, we built our shutter glasses based on a novel double-cell shutter design, which consists of two layers of differently configured regular LC shutters. The first layer is a regularly cross-polarized LC shutter (normally white (NW)), which is transparent if no voltage is applied. The second layer has equally oriented polarization filters on both sides and thus it is opaque (normally black (NB)) if no voltage is applied. This combination of shutters functions so that the NB shutter opens quickly while the NW shutter closes quickly. These shutters are ideally suited for an asymmetric use case: our shutters need to be open for only $1/360th$ of a second and closed for $5/360th$ of a second. During the longer closing time, both shutters relax one after the other: first, the NB shutter closes fully and then the NW shutter opens completely (Figure 10.5). In addition, using a stack of two shutters improves the contrast ratio, an important property in the context of our system.

In a six-user stereo-projection system each individual shutter must blank 11 of 12 displayed images. For a left eye shutter of a particular user three distinct cases can be considered (similarly for a right eye shutter):

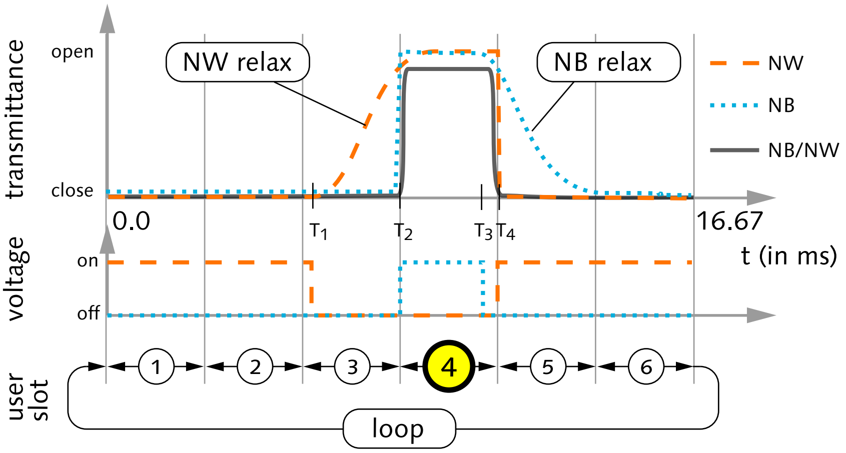


Figure 10.5: *Illustration of the double shutter functionality (top), the electrical shutter driving pattern (middle) and the time slots for each user (bottom). At 60Hz a 16.67ms time frame is divided into six adjacent user time slots of equal length. The diagram shows the timings of the double cell shutter of user 4, who receives an image during the fourth slot lasting from* T_2*=8.34ms to* T_4*=11.12ms. The NW shutter is switched off at* T_1 *about 2ms before* T_2 *to ensure its relaxation and thus maximum light transmission at the beginning of the following opening period. The NB shutter is still blocking light during this NW relaxation phase and immediately opens when the voltage is applied at the beginning of the 4th time slot (*T_2*). At the end of the opening period of 2.78ms (*T_4*) the NW shutter is immediately blocking the light transmission as the voltage is applied. The NB shutter is switched off for relaxation slightly before (*T_3*=11ms).*

1. The user's right eye image is separated by polarization.
2. The left eye images of the other 5 users are blocked by the shutter operation.
3. The right eye images of the other 5 users are blocked by polarization and shuttering.

The first case contributes only the relatively low crosstalk of standard polarization-based systems. The second case is addressed by our new double-cell shutter design, which provides fast switching times and high contrast to avoid crosstalk. In our setup the shutters in closed state must block five times more light as compared to the case of active stereo displays. Double-cell shutters help with this requirement since the total contrast ratio is the product of the contrast ratios of the NW cell and the NB cell. The third case contributes at least one order of magnitude less crosstalk than the other two cases since the light is blocked by shuttering and polarization.

We designed our wireless shutter glasses (Figure 10.6) such that their principal state can be controlled from the application, independent of the basic clocking. The communication to the shutter glass controller is realized by using the µracoli implementation [344] of the two lower levels of the IEEE 802.15.4 protocol stack for wireless personal area networks (which form the basis of the Zigbee protocol). There are two different aspects that can be programmed:

- The general assignment to one or more of the six-user time slots. This control can be used to implement a VIP (Very Important Person) mode by assigning two or more time slots to a single person. We often have the case that the system is used by less than six individuals and thus we use this control to increase the brightness by assigning more than one time slot to one or more users.
- A transition from shutter mode to full-open mode and vice versa. In regular operation the shutters are open for only $1/6th$ of the time and thus everything but the display is perceived as quite dark. However, if six people are in front of the display discussing various aspects of their application, it quite often happens that they look at each other or do not look at the display at all. They may even move to a whiteboard to continue discussion. In these cases we open the glasses and turn off the shutter mode using simple heuristics based on the head tracking information.

There are many other uses for application-controlled shutter glasses. Particularly in multi-display environments (e.g. [270], [175]), where users interact with a variety of displays, shutters need to sync to the currently faced display and should be turned off if it is a 2D display or only 2D content is presented.

Figure 10.6: *Our custom shutter glasses consist of two double cell shutters, a Zigbee radio module, a rechargeable lithium-polymer battery and the shutter driving circuit. The housing also contains multiple threaded holes for assembling different IR-reflective marker configurations.*

10.5 Augmented Group Navigation

Projection-based multi-viewer systems expose each user to an individual pair of images, which can be computed in a way that the following two properties are ensured:

- *Perspectively-correct perception of the virtual world by each user*. The perspective projection is defined by the relationship between the physical position of a user's eye and the physical position and size of the projection screen. Since the physical position is different for each user, the perspective projections and the computed images are also different. The correct perspective enables correct size, shape and distance perception in the virtual world. Regular stereoscopic displays present the same stereoscopic image pair to all users. Since the perspective can only be correct for at most one user, all the other users perceive distorted versions of the virtual objects [4].
- *Perception of a consistent virtual world among all users*. The users and the projection screen are placed in the virtual world in exactly the same spatial configuration as in the real world, apart from a global scaling factor. Only this configuration ensures that the virtual world is consistently perceived by all users (e.g. they are seeing a virtual model as if they were standing around a real version of the same model). Bare-handed pointing becomes possible and each user interprets the pointing gesture as in the real world, accompanied by some limitations with respect to accuracy [297].

These two properties enable a group of co-located users to visually perceive a virtual model as if it were a real part of their shared physical environment.

Navigation is a basic interaction capability of almost all virtual reality applications. However, the availability of projection-based multi-viewer systems introduces a new problem for the navigation through a virtual world, which did not exist in classical VR systems with a single head-tracked user. In such single-user systems, all the users share the same perspective and if the navigator is moving along a collision-free path, all the group members also perceive the path as being collision-free. In multi-viewer systems all of the users are simultaneously moved around along with a virtual representation of the projection screen. Since projection displays can be quite large (e.g. our display is 4.3m wide), the users are typically distributed around the space in front of the display. In this configuration, they may not fit through constrictions such as doors or aisles even if the navigating person chooses a collision-free path. The co-travelers might be passing through walls, which can be annoying and irritating and the navigator may not even be aware of these problems.

Unfortunately, there is no general solution to this problem, which maintains a correct perspective *and* the consistency of the shared virtual world for all users without requiring that all users are looking from the same position as the navigator, which

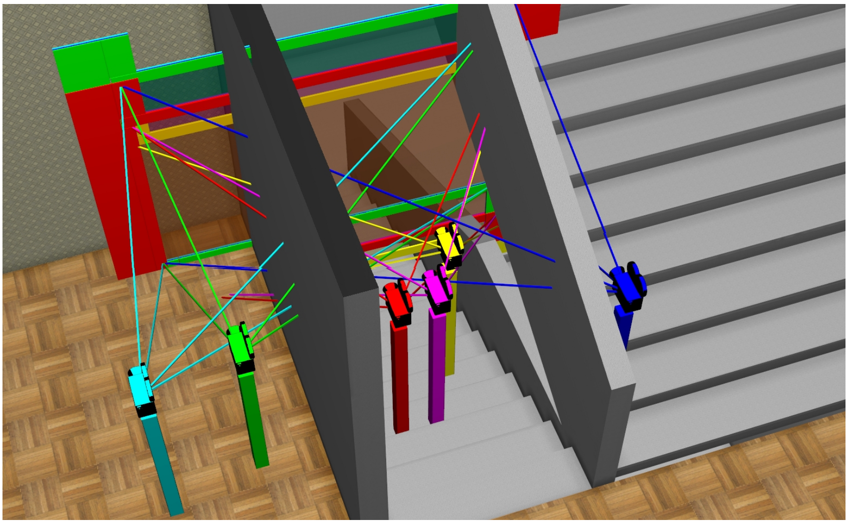


Figure 10.7: *This debug view of our system shows six users (illustrated as colored columns with a camera on top) moving together towards stairs in a virtual building. Only the two users standing very close to the navigator (red) follow on the chosen path to a lower floor. Two miss the entry to the staircase (cyan and green), while another user (blue) is moving at the the neighboring stairs upwards. This undesirable dissociation of the group experience is the geometrically consistent result of moving in the same virtual direction while maintaining the physical configuration in front of the shared display.*

is physically impossible. However, we suggest various approaches to mitigate the problem:

- *Stop and crowd*. As a simple solution we perform collision detection for each user and stop the navigation if at least one user collides. To continue, users need to resolve the collision (e.g. by moving closer to the navigating user).

- *Distort*. On a path towards an obstacle, the head position of the user is moved towards the head position of the navigator until the collision is avoided (Figure 10.8). Changing the head position incurs a distortion of the perspective and thus the surrounding space is no longer correctly perceived. In an extreme case, the system collapses to a two-view system and all the users see the same images, which still results in a different perception of the size, shape and position of virtual objects.

- *Detour*. On a path towards an obstacle the system interferes and moves the user along a collision-free path while maintaining a perspectively correct rendering (Figure 10.9). This mode has three degrees of freedom if a planar movement is assumed. The user along with the screen representation can be translated in two dimensions as well as rotated around the up-axis. We use a heuristic to determine the translation, which moves the user towards the navigator as much as necessary to avoid a collision. The rotation is ignored up to this point. This approach temporarily gives up the consistency of the shared virtual world, while the perspectively correct perception is maintained.

- *Fade*. If a user is on a path towards an obstacle, this object is slowly faded out during approach. The opacity of the obstacle is defined as a function of the distance to the respective viewpoint and the current velocity of the user. In our current implementation object geometry less than four seconds away starts fading out such that it is invisible one second before the linearly predicted collision. As a result, the irritating collision with the obstacle is avoided while users may still examine details of an object when standing in close proximity. However, during travel the consistency of the shared model is affected since other users may see an unaltered object.

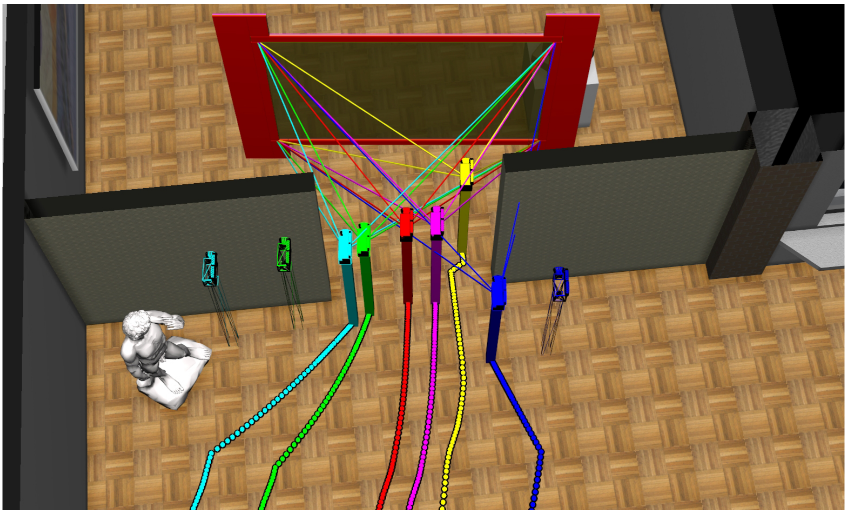


Figure 10.8: *Distort: The viewpoint position (opaque representations) is shifted towards the open passage if the physical head position of a user (drawn as wireframe) would collide on its original course. Users are represented as colored icons. Their viewing frustum is defined by their position and the virtual screen representation. Shifting the head position away from the correct position leads to a distorted perception of the virtual world.*

The last three approaches return to an artifact-free rendering if there are no more constrictions. They use smooth transitions in and out of these modes to avoid disorientation. *Stop and crowd* and *Fade* depend only on each user's individual viewpoint position and the motion velocity. *Detour* and the *Distort* redirect a user from a given navigation path. In these cases we assume that the navigating user chooses or is forced along a collision-free path. For all other users we use ray-object intersection tests to find obstacles in the steering direction. In case that intersections closer

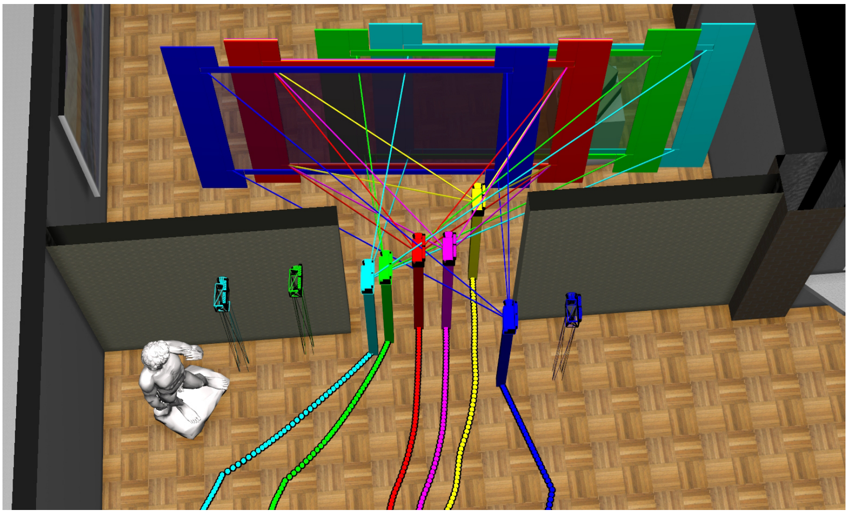


Figure 10.9: Detour: The viewing frustum associated with a particular user is shifted towards the open passage if the physical head position of the user would lead to a collision. Users no longer observe the virtual world through the same window.

than a velocity-dependent distance threshold are found an obstacle-avoiding movement toward the path of the navigating user is added to the prevailing viewpoint motion. The distance threshold is a function of the steering velocity. The velocity of the obstacle-avoiding movement is defined as a function of the distance to the main navigation path and the current steering velocity ensuring that it is zero once the predicted collision would occur. In our current implementation we compute the distance threshold such it considers objects up to five seconds ahead (e.g. 3.5m at a velocity of 0.7m/s).

As a consequence of these viewpoint adaptations we must deal with two representations of each user's position: one is corresponding to the original head position defined by the tracking system and the other corresponds to the adjusted head point position for collision avoidance (Figures 10.8 and 10.9). Once looking ahead from the adjusted head position does not predict any more collisions, we do not move the user further toward the path of the navigator. As a result the collision-avoiding motion tries to minimize the deviation from the original path as can be seen in Figure 10.9. The adjusted head position can be moved back to the physical head position if the ray-object intersection tests for the original head position do not predict any further collisions and if there is no obstacle between the original head position and the adjusted head position.

This heuristic approach for collision avoidance has obviously many limitations including the shape and orientation of obstacles that can be robustly detected to avoid collisions. However, in practice it works well in typical use cases such as an architectural walk through or a tour through a virtual museum. In particular, it was sufficient for our purpose of evaluating the general applicability and user acceptance of these different collision handling techniques for group navigation.

The *Detour* and *Distort* techniques can be compared to the concept of redirected walking [247,279,322], since they redirect a user from a given navigation path. Redirected walking aims to extend the range of physical walking motions in virtual environments by redirecting the user from an actively controlled walking path. In our system a group of users is being navigated by an operator as if being passengers in a vehicle that is driven by somebody else. The path of the driver is unaffected by the redirection techniques. Thus the passive passengers do not have a particular expectation of a movement through the virtual environment or a proprioceptive reference that would allow them to judge the extent of redirection. The only reference are the positions of the other users in front of the display and the size of the constriction in the virtual world. This makes it very difficult to even detect the redirection.

The described augmented group navigation techniques are all orthogonal to the actual navigation technique being used by the navigating user. In our demonstrations and user studies we use predefined navigation paths or direct steering techniques. As a result of unsatisfactory experiments with handheld steering devices we developed a stationary group navigation device, the Spheron (Figure 10.10). The most prominent feature of the Spheron is a large 3D trackball for rotating the view. An elastic handle is mounted at the base of the device for controlling the movement through the scene. The Spheron is centrally placed within the shared action space and is easily accessible by all users so that taking turns while navigating is easily achieved. The physical presence of the device itself and the required manual operation foster mutual awareness among the group of the interactions performed by the navigator and allow others to quickly take over control and interfere if desired.

The Spheron is inspired by the CAT [123], which was also developed for group interaction in large-screen environments. Compared to the CAT the Spheron has a stronger affordance for rotation through the 3D trackball. Both, examining objects from different sides and looking around in the environment are particularly necessary in the context of group navigation using a single wall-sized display. As another important difference to the CAT the trackball can be equally well accessed from all directions whereas the steering wheel of the CAT can be awkward to grasp if it is tilted. The 3D trackball of the Spheron always remains at its position. It can even be accessed blindly and thus facilitates frequent changes of the operator among a group of users.

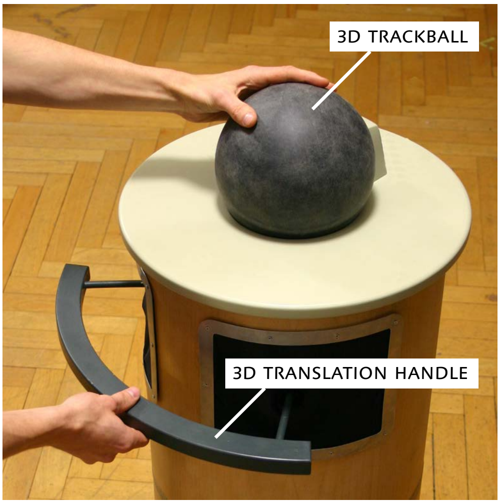


Figure 10.10: *The Spheron group navigation device. Moving the elastically suspended handle induces a corresponding rate-controlled motion through the scene. Rotation of the 3D trackball results in changing the view direction. Alternatively, for viewing a single model, the trackball rotations can be mapped to rotations of the model.*

10.6 Results and Evaluation

We set up the six-user system in our lab with a rear-projection onto a 4.3m by 2.7m screen (Figure 10.1). The system is driven by two computers in a master-client architecture. All application logic is computed on the master node and the state updates are distributed over the network to the clients. The client applications render the scene for the individual users based on head tracking information directly fed to the clients. Using this setup we typically achieve application frame rates above 30Hz for the scenarios in our user study. The client applications run on a single HP Z800 computer equipped with two Intel Xeon X5680 six-core processors running at 3.33GHz, 96GiB of main memory and three NVIDIA Quadro Plex 7000 graphics subsystems. The system configuration of the rendering computer can be seen in Figure 10.3. Our demonstrators are based on the free software AVANGO^{NG} [15, 184] under 64 bit Ubuntu 9.04. The end-to-end latency from tracking the user's motion input to the display update is about 80-120ms depending on the actual rendering frame rate. This is very similar to a single-user system, since the actual rendering threads run in parallel on separate cores and GPUs.

We use a large-field optical tracking system for tracking the head positions of the users and further input devices. For the image separation we currently use different prototypes of shutter glasses. Three of them consist of FLC shutters, the other three are based on the novel LC double-cell design.

10.6.1 Technical Evaluation

We used a Spectroradiometer (Konica-Minolta CS-1000A) to measure the relative impact of the display components on the overall optical quality. We were particularly interested in the perceived brightness at the user's eye behind the shutter glasses and the potential crosstalk from images of the other users. We also wanted to verify that adding polarization filters in front of the DLP projectors only slightly decreases the brightness level per user. Our measurements show that it is in fact only decreased by 12% since the shutter glasses are polarizing the light anyway. We found that the brightness linearly increases with the length of the opening periods of the shutters. This behavior verifies that each user receives $1/6th$ of the total brightness of the six projectors, which is equivalent to the brightness of a single projector. These tests were done with FLC shutters and double-cell shutters with precisely adjusted timings, which resulted in similar behavior.

As explained earlier, ghosting images are much more an issue in multi-user projection systems than in common single-user stereo displays. In particular, ghosting of images from other users are more noticeable than left-right eye ghosting since the other users move independently of oneself. Thus the ghosts move within ones images, which makes them more salient. To avoid ghosting we developed the double-cell shutter glasses, which provide fast switching times and a high contrast ratio. An informal comparison of the novel double-cell LC-shutter glasses to the FLC-shutter glasses indicates that double-cell shutters eliminate the ghosting to a non-perceptible level in most scenarios while FLC-shutters show slight but perceptible leakage in dark image areas, which is not visible with the double-cell shutters.

10.6.2 User Feedback

In general we observe that people are more enthusiastic about exploring a virtual environment as part of a group with perspectively correct views for each user than in a regular stereoscopic environment. Even if the displayed content does not directly correspond to their interest, it becomes a relevant part of their shared reality through the immediate exchange with others. The Spheron turned out to be an easy to use navigation device, which does not require any explanations in most cases. For evaluating our augmented group navigation techniques we performed a pilot study followed by a formal user study.

Pilot Study

For the first study we invited two groups of five users. One group consisted of students of industrial design, the other group of students and alumni from the department of architecture. We introduced the five participants of each group to the technology and advised them to notify us of any problems such as excessive crosstalk, loss of orientation or nausea. An instructor also wearing shutter glasses steered the group through the model of a museum with several rooms and exhibits using the Spheron device. The different group navigation techniques were presented in a predefined sequence starting with the *Stop and crowd* mode.

For the evaluation we chose a semi-structured group interview, allowing us to introduce new questions during the discussion. We were interested in two main topics: the overall usability of the system and, more specifically, the users' experience with the different augmented group navigation techniques. The different techniques were rated on five-point Likert scale. Besides an overall very positive assessment of the system the main observations of the pilot study are:

- Ghosting was not explicitly reported as a problem. Nevertheless, the participants remarked that they generally preferred the double-cell shutter glasses over the FLC-glasses for reasons of image quality.
- Everyone agreed that bumping through opaque walls is not acceptable and augmented group navigation techniques are needed.
- The *Distort* technique was consistently rated low for two major reasons: the distorted geometry breaks the realistic impression of space, and the animated transitions in and out of the *Distort* mode can contribute to a feeling of dizziness.
- The *Stop and crowd* solution provides the most realistic experience. However, it quickly becomes annoying if the group needs to pass through a number of doors and aisles.
- The opinions on the *Fade* mode varied. Some liked it for being a convenient option to maintain one's relative position in the group, while others found that it would deteriorate the appearance of the model.
- The architects preferred *Detour* as it provided the most realistic experience of the virtual building without the necessity of crowding.

Statistical analysis of the user ratings for the different techniques using the Kruskal-Wallis test for ordinal data and the Mann-Whitney U test for post-hoc comparisons revealed a significant preference for the *Detour* technique over *Distort* ($p < 0.05$). The other reported preferences did not become significant, but there are clear trends visible in the data. The strong preference for *Detour* as compared to *Distort* demonstrates the users' aversion to perspective distortions, which was the main reason to develop the multi-user projection system.

10.6.3 User Study

Overall the pilot study revealed that *Detour* and *Fade* were considered the most useful augmented group navigation techniques. We compared these two techniques in a controlled user study to evaluate their usability in relation to the baseline of untreated collisions during group navigation.

Experimental Setup

We modeled a simple virtual architecture containing full-size 3D models of cars and photos of their real counterparts (Figure 10.11). Navigation was automated and took the participants with a fixed velocity of 0.7m/s on a predefined path through the environment. The automated tour eliminated any bias of an operator steering in a slightly different way in each test condition. It also allowed testing with groups of six users. During the tests the users could move freely in front of the screen to assume desired viewpoints.

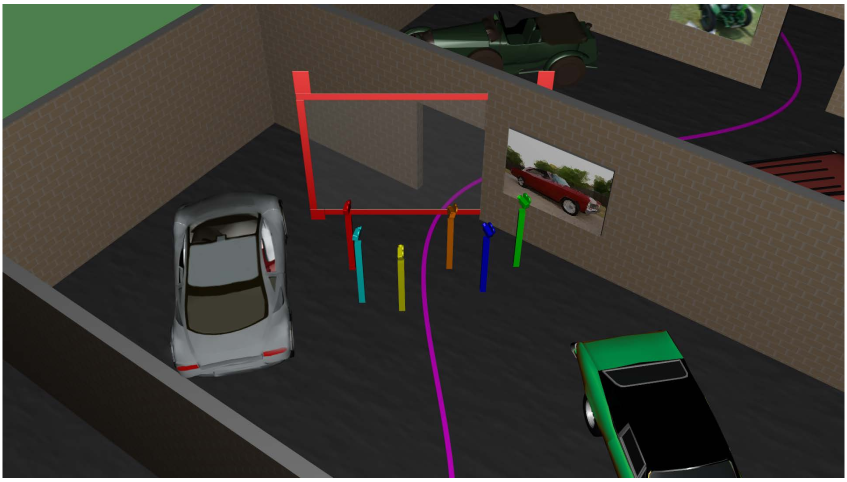


Figure 10.11: *Visual search task: The group is automatically moved along a given path through a virtual car exhibition. Participants were asked to search for a set of given features on the 3D car models or in the exhibited pictures.*

The participants were asked to perform a search task in the virtual environment. Each user received a sheet of paper showing 15 pictures with details from virtual car models and their real counterparts. We informed them that exactly seven of these details could be found in the presented scene. The users' task was to memorize the pictures and search for the respective details during the automated tour with one of both group navigation techniques applied. Thereafter, recognized details should be marked on the paper. During a repetition of the same tour everybody got the chance to verify the findings, while another group navigation technique was applied. This primary task ensured that they would concentrate on the application content rather than focusing for the perceptual artifacts of the tested group navigation technique.

Besides the factor *navigation technique* we included *information exchange* as another independent variable. We were interested to see if solving the memory task individually in the *single* condition or exchanging information within the *group* would affect the discovery. While we did not permit gesturing or talking in the *single* condition, mutual information exchange during the tour was encouraged in the *group* condition. In both conditions everybody noted the recognized detail individually again. Our hypothesis was that individual users would perform better if they are able to exchange information about recognized details in the virtual environment. We used two different sets of models and pictures with comparable difficulty in both conditions.

Participants

Seven female and 17 male users aged between 20 and 30 years participated in this study. None of them reported problems with stereoscopic vision. Six of them claimed to have extensive experience with interactive 3D computer graphics while three reported to have no prior experience with the technology. All 24 participants were university students.

Design and Procedure

First, we introduced the participants of each group to the general characteristics of our multi-user VR system. The interior of a car was shown and the participants were asked to touch items like the rear view mirror or the steering wheel. As everybody could see what the others were doing, it became clear that the displayed virtual environment was consistent for all involved users. We also introduced the mentioned collision problem for multi-user navigation. The automated navigation was started without any collision handling. We made sure that the users were changing their position in front of the screen such that everybody experienced the effect of bumping through virtual walls. Directly after this experience we asked the participants to rate the severity of the collision issue on a five-point Likert scale (ranging from -2 (unacceptable) to $+2$ (very good)). We also asked whether anybody felt symptoms of cybersickness during this training tour, offering to abort the study if necessary.

Both augmented navigation techniques were presented two times to each participant, once for each condition of the factor *information exchange*. The order of all conditions was equally distributed among the four test groups using a reduced Latin square design.

Results

Users strongly complained about bumping through walls and consistently rated this baseline condition negatively with a mean of -0.5 ($sd = 0.71$) while both augmented group navigation techniques were rated positively with a mean of $+0.5$ for *Detour* ($sd = 1.1$) and $+0.7$ for *Fade* ($sd = 0.71$) (Figure 10.12). Statistical analysis using the Kruskal Wallis test and the Mann-Whitney U test for post-hoc comparisons revealed this difference to be highly significant ($p < .01$). There was no statistically significant difference between the ratings for *Detour* and *Fade*. We observed, however, that 22 of 24 users have a clear preference for one of both techniques. The 10 participants voting for *Detour* explained their choice with a more realistic appearance of the architecture and many among them claimed that they did not even realize the redirection. *Fade* instead was preferred by 12 participants who identified the consistency of the virtual view to their physical position in the group as the most relevant feature of our multi-user VR system.

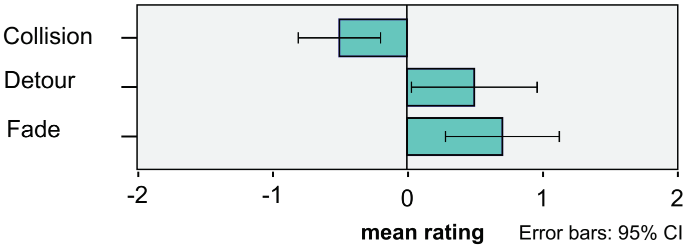


Figure 10.12: *Mean ratings and confidence intervals for the three tested group navigation techniques.*

We also found evidence that users can indeed benefit from immediate communication within our virtual environment. The success ratio for recognizing details in the virtual environment was about nearly twice as high in the *Group* condition ($mean = 0.43$, $sd = 0.16$) as compared to the *Single* case ($mean = 0.25$, $sd = 0.22$). This statistically significant result ($T_{(46)} = 3.437, p < 0.05$) was expected. It indicates that our multi-user VR system indeed enables users to exploit common advantages of collaboration.

10.7 Conclusions and Future Work

We designed and implemented the first large screen stereoscopic multi-viewer display for six tracked users, which provides precise horizontal and vertical parallax. The system runs at 360Hz, which results in 60Hz per user. Left and right eye separation is provided by polarization. This efficient combination of active and passive stereo achieves almost the same brightness per user as a single-user active stereo system based on the same type of projectors would. Our intelligent shutter glasses enable application-level control of the shutter timing for better usability and higher brightness if less than six users use the system. Novel augmented group navigation techniques facilitate the exploration of environments with constricted navigation paths by avoiding collisions of fellow travelers.

Our augmented group navigation can be further refined in many directions. In particular, formulating the combination of different collision-avoiding strategies as an optimization problem could result in a good heuristic to select the best strategy for each situation and provide smooth transitions in between. So far we have focused on keeping the shared space intact as much as possible while navigating. While this seems highly desirable in most cases, there might also be applications where it is beneficial to temporarily relax this requirement. Each user might stroll through a museum on an individual path, but has the possibility to rejoin the group if desired. Tele-collaboration systems such as the blue-c [112] have only been developed for a single tracked user. Our approach would enable immersive group-to-group collaboration. However, navigating a group consisting of local and remote users in a consistent and transparent way remains a considerable challenge.

Our six projector array is currently a unique prototype system. The integration of three separate projectors into a three-chip DLP projector is the next step forward. This would reduce the complexity of setting up and running such a system. In addition, it is much more energy efficient and produces much less heat considering a three-chip DLP projector generates the different primary colors by splitting up the white light beam from the projector lamp instead of using color filters. Our approach of turning a regular monoscopic DLP projector into a stereoscopic projector by switching between two input and formatter boards could be applied twice, and thus we could switch between four input streams. While this may come with the disadvantage of losing color depth and reduced brightness, it would allow us to build an active stereo two-user single-chip color wheel projector, an active stereo six-user three-chip DLP projector or a twelve-user active-passive system consisting of two such three-chip projectors. With only minor modifications, all of these systems can be built from readily available components.

Multi-viewer projectors would significantly improve the usability of various immersive system designs, which typically involve groups of three to six users. Multiple users could gather around a table-top display like the Responsive Workbench [183]. Since the invention of the CAVE [65], the original design has been considerably improved by using more walls and higher resolution (e.g. the C6 [351]). We are looking forward to seeing our technology serve as an instrumental part in the evolution towards the first C6x6.

Acknowledgments

This work was supported in part by the European Union under grant 217140 (project IMVIS), the German Federal Ministry of Education and Research (BMBF) under grant 03IP704 (project Intelligentes Lernen) and the Thuringian Ministry of Education and Cultural Affairs (TKM) under grant B514-08028 (project Visual Analytics in Engineering). We thank our IMVIS partners LC-TEC DISPLAYS AB, Personal Space Technologies and Centro Ricerche FIAT for the excellent collaboration throughout the project, Frank Pudlowsky and Jörg Krall from NVIDIA's Professional Solutions Group as well as Joachim Girke from Hewlett-Packard GmbH for providing extraordinary support with our complex hardware setup, and the members and students of the Virtual Reality Systems group at Bauhaus-Universität Weimar for their help with getting everything to work. The model of the David statue in Figure 10.1 is courtesy of Stanford's Digital Michelangelo Project.

Part III

Further Examples and Conclusions

Chapter 11

Case studies

With the experiments and technical developments described in Part II of this thesis we pursued three very different approaches to cooperative user interfaces. We explored potential benefits of bimanual cooperation in desktop-based CAD applications, we designed a novel multitouch technique based on knowledge about bimanual cooperation behavior, and we developed a novel display technology for multi-user cooperation in immersive 3D environments. Part III of this thesis aims to illustrate the design principles of *workspace coherence*, *complementary capabilities*, and *emergent territoriality* (see Chapter 6) with experimental applications and further technical developments in the realm of collaborative virtual reality.

Sections 11.1, 11.2.2, and 11.3.1 report on joint work with Stephan Beck, André Kunert, and Bernd Fröhlich. Section 11.1 summarizes parts of a publication at IEEE VR 2013 with the title "Group-to-Group Telepresence" [26]. Section 11.2.2 and Section 11.3.1 refer to research and developments that were published at ACM CSCW 2014 under the title "Photoportals – Shared References in Space and Time" [189]. Section 11.2.1 reports on joint work with Ferran Argelaguet, André Kunert, and Bernd Fröhlich. It summarizes parts of a publication at IEEE 3DUI 2010 with the title "Improving Co-located Collaboration with Show-Through Techniques" [8]. Please refer to the full papers for detailed descriptions of the technologies and experiments. The experimental assembly design application described in Section 11.3.2 was developed by Sebastian Utzig as part of his bachelor thesis at Bauhaus-Universität Weimar.

11.1 Workspace Coherence in Remote Collaboration

The multi-view 3D projection system described in Chapter 10 enabled the joint experience of immersive virtual environments for up to six collocated users. Our observations during frequent public demonstrations of the system confirmed the expected relevance of immediate exchange between users about the displayed environments. In particular, we observed that gestural communication like pointing and tracing was used a lot to indicate interesting details and confirm the jointly perceived presence of virtual objects. We also saw people posing with their bodies in meaningful spatial relations to the virtual environments. For example, they simulated sitting in a virtual car model with their hands at the steering wheel, or they gauged model dimensions like the width of a door with their arms.

However, the vertical display only supported the collaboration of collocated users in a side-by-side configuration. Interacting with the virtual content was impossible in a face-to-face setting. With respect to the increasing importance of computer applications for remote communication, we considered the involvement of remote participants through virtual avatars. Commodity depth-sensing cameras were emerging at that time, and the feasibility of creating photorealistic 3D video avatars on this basis had been demonstrated [218].

We realized a similar setup for real-time 3D capturing, data transmission, and reconstruction. The novel technology in conjunction with two multi-user 3D screens allowed us to build the world's first group-to-group telepresence system [26]. Two remote groups could meet in a coherent virtual environment. One of these setups supported two collocated users, the other one allowed to involve up to six users. The users could directly perceive each other's presence, identity, and activities (Figure 11.1). Furthermore, we developed a set of appropriate interaction techniques, including a shared "world in miniature" (WIM [325]) to support their joint navigation through larger virtual spaces.

Visitors of our lab were generally excited about the novel possibilities of this group-to-group telepresence system. Similarly to the above mentioned observations with collocated multi-user virtual reality, we saw users communicating with remote participants using body language and pointing gestures. To verify the positive feedback, we performed a formal study with four groups of three users. The study involved three subsequent phases of using the system for about 45 minutes in total. During the welcome and introduction session, the group of participants was asked to greet two remotely present experimenters with virtual handshakes (visual feedback only, see Figure 11.1). In the second phase, we tested the comprehensibility of pointing gestures from local and remote users at individual buildings of a miniature city model (Figure 11.3). In the third phase, both groups negotiated a joint tour through the virtual city and then followed each other on the chosen path through the model at original scale (Figure 11.2).

Figure 11.1: *Two groups of users meet in a virtual city model. The image to the left shows the situation at a six-user display. The same situation on the other side with a two-user display can be seen on the right. Figure reprinted from [26] © 2013 IEEE*

Figure 11.2: *Group navigation in a telepresence setting. Left: Two groups use a world in miniature to find their position in the virtual environment. Right: One group follows the other during a virtual city tour. Figures reprinted from [26] © 2013 IEEE*

11.1.1 The Effectiveness of Gestural Communication

The comprehensibility of pointing gestures was tested in two conditions: local and remote. In the local condition, the experimenter shared the same collocated setup as the participants. In the remote condition, the experimenter used the other system and the participants saw his 3D video avatar pointing at the scene (Figure 11.3). The task required each user to identify in each condition ten townhouse models of about 2 x 2 x 2 cm that were indicated through pointing, either of by a local user or the 3D video avatar of a remote person.

During informal use of the system, we did not observe immediate communication issues. Our formal pointing study, however, revealed limitations of perceived accu-

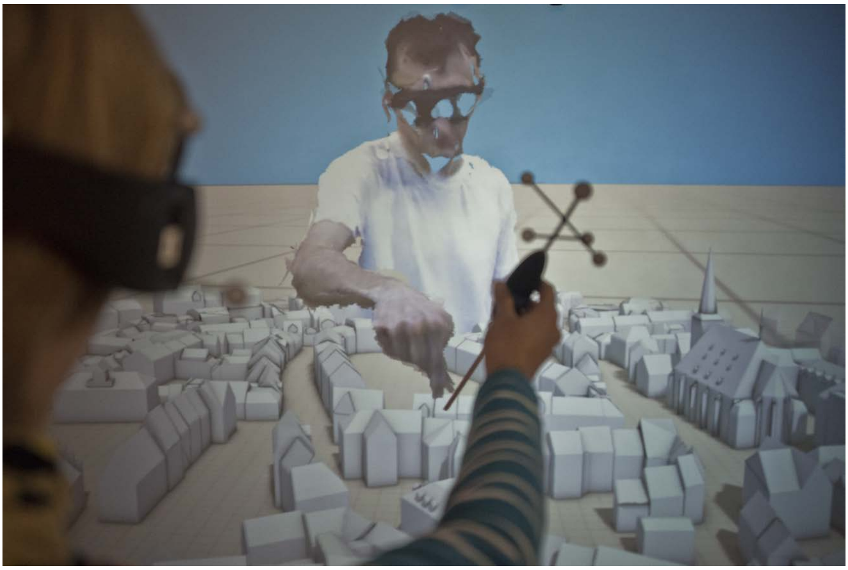


Figure 11.3: *A remote user pointing at a building and a local user identifying it during the pointing study. Figures reprinted from [26] © 2013 IEEE*

racy in both conditions. Two of the four groups made one identification error in the remote setting. Given that each participant of each group was asked to identify ten different buildings, this sums up to a total error rate of 1.67%. As a reason for the misunderstanding the participants noted visual noise of the avatar's finger. Interestingly, one group also made two identification errors in the collocated condition. We speculate, that depth perception was hampered due to conflicting cues of vergence and accommodation (cf. [297]).

11.1.2 Subjective Workspace Awareness

Subjective participant feedback was captured using a rating scale consisting of 10 topics related to the overall impression of the system, spatial perception, communication support, perceived co-presence, and opinions on system details. Each topic was covered by several separate questions that were evaluated in conjunction. The questionnaire was and adapted and extended version of that used by Mühlbach et al. in their study on stereoscopy and eye contact in videocommunication [239].

The overall feedback on the system was very positive and participants expressed their interest in using such a system frequently (Figure 11.4 dark blue). The 3D video avatars and the virtual city model conveyed a strong sense of spatiality (Figure 11.4 pink). The users reported that group coordination was effective, but they also reflected the system's limitation to support communication through gestures and gaze with correspondingly lower ratings (Figure 11.4 green). Consequently, the illusion of physical co-presence with remote participants was rated relatively low (Figure 11.4 orange). Clearly, our 3D video avatars were of limited quality (Figure 11.1) and could not compete with the actual physical presence of the collocated participants.

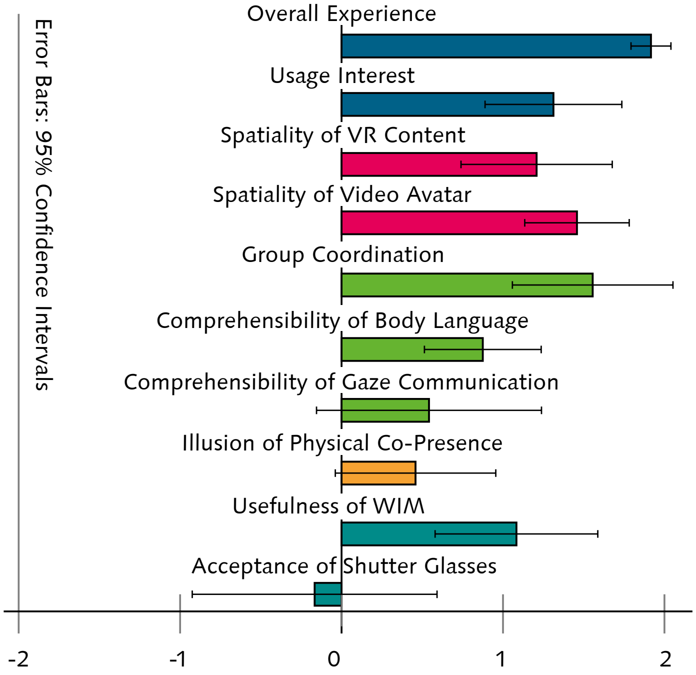


Figure 11.4: *Average user ratings of system characteristics, clustered by similarity of topics. Blue bars show the participants' general feedback and interest. Pink bars show the subjective quality of spatial perception. Green bars indicate aspects of group communication. Yellow bars show the perceived sense of the avatars' copresence. The turquoise bars at the bottom relate to functional and ergonomic aspects of interface. Figure reprinted from [26] © 2013 IEEE*

11.2 Territoriality in Immersive Virtual Environments

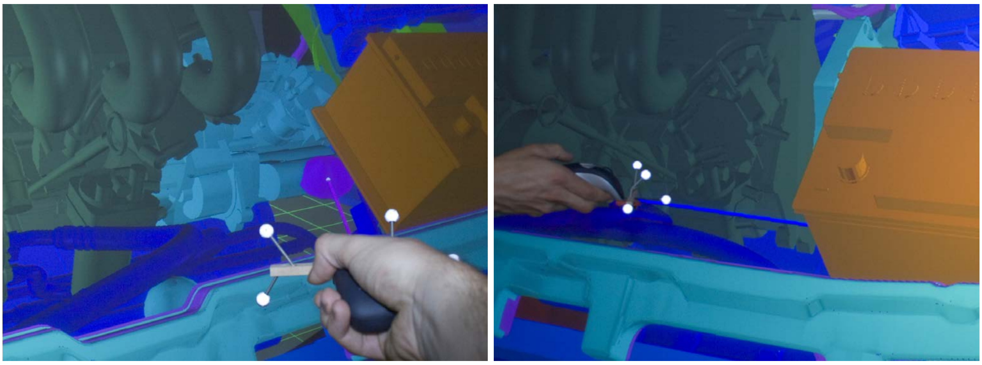
Territoriality is an emergent social behavior. Scott et al. showed that collaborating people spontaneously partition their shared space if it is large enough and provides the appropriate means for parallel activities [311]. Collaborative computer applications, therefore, primarily need to provide a suitable interaction space, e.g., through large shared displays (e.g. [294, 295, 328]). This allows users to find individual positions, from which to approach the application content.

In case of immersive 3D content, the user's position also defines what can be seen. If all users share a single view of one tracked "group leader", the secondary viewers cannot explore the 3D visualizations individually, but they experience a distorted and unstable virtual environment. Therefore, each user of a collaborative virtual reality system must be provided with individual images that represent the scene from their respective viewpoints. This fundamental feature of our multi-user 3D projection system, described in Chapter 10, enables joint perception in a shared and consistent 3D environment. Since the users are provided with personal views, they can also separate their activities from the group by focusing at different parts of the shared environment. Rendering separate images per user, furthermore, enables specialized views according to different roles or to solve situational conflicts [4]. Multi-view display technology, thus, enables *workspace coherence* as well as *territoriality*.

Chapter 10 described a situation where inconsistencies of the physical interaction space and the shared virtual environment can lead to unclear joint affordances. People in front of a large screen generally position themselves laterally to each other with appropriate interpersonal distances (see proxemics in Section 6.3). Standing behind each other is not a suitable configuration, since they would obstruct each other's view towards the display surface. Over time, they establish these locations as their private territories, from which they start their interactions with others and the application content. If a group of people is, instead, walking through an environment with narrow passages, e.g., a building with doorways, they tend to walk one after another to fit through the constrictions while maintaining appropriate distances. During a virtual architectural walkthrough with a large multi-user 3D display, both affordances are in conflict, with the constraints of the physical setup being dominant. We resolved this issue with augmented group navigation techniques, that accomplish collision-free virtual navigation through temporary deviations of the virtual viewpoints from the physical user positions (see Chapter 10). Slight deviations from spatial consistency and physical realism can facilitate the collaborative exploration of virtual environments.

11.2.1 Visibility in Dense 3D Environments

The appearance of 3D objects and environments clearly depends on perspective. Artifacts and features that are visible for one person, may be completely or partially hidden by surrounding geometry from the viewpoints of others (Figure 11.5). To see an object someone else is referring to, it may be necessary to obtain a very similar viewpoint, i.e., to look over the other's shoulder. This situation occurs primarily in densely packed environments like a warehouse or an engine compartment. However, looking over one's shoulder is often an undesired violation of proxemic behavior. In formal presentations, people prefer to maintain distances of more than one meter [125].



(a) Presenter view (b) Observer view

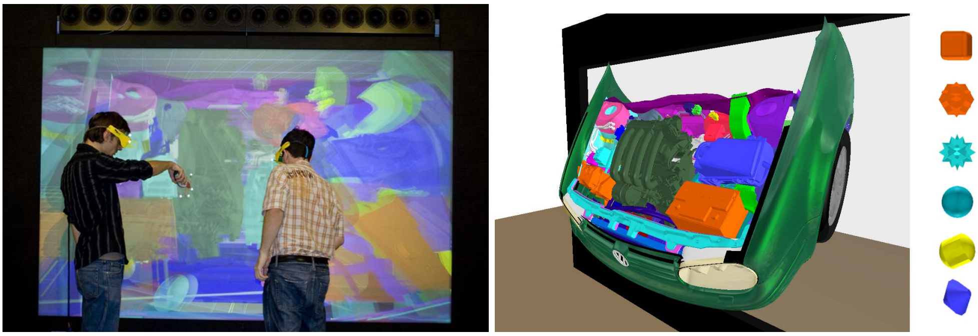
Figure 11.5: The issue of interpersonal occlusion. The image on the left shows a scene the perspective of a presenter pointing at an object of interest. The image on the right shows the same situation from the perspective of an observer, who cannot see the indicated object. Figures reprinted from [8] © 2010 IEEE

(a) Cut-away view (b) Fade-away view

Figure 11.6: The images show the same situation as the observer view in Figure 11.5 but with show-through techniques. Objects hidden behind other geometry become visible through cut-away (left) or fade-away visualization (right). Figures reprinted from [8] © 2010 IEEE

In collaborative virtual environments the problem can be alleviated. The visibility or transparency of occluding geometry can be dynamically changed such that the selected features *show through* (Figure 11.6). We studied the effects of such augmented viewing techniques on proxemic behavior and spatial understanding. Our hypothesis was that users would maintain more comfortable distances between each other. On the other hand, we also expected that the ability to later retrieve a number of shown objects could be higher, if the users had to actively find an appropriate viewpoint to be able to see them during their initial presentation.

The experimental task consisted of two phases. During the presentation phase, the experimenter was pointing at target objects in a virtual car engine compartment (Figure 11.7). He always kept pointing until the participant confirmed their recognition of the shown object and its location. Thereafter, the participant was asked to retrieve these objects in the scene based on memorized information. In two *show-through* conditions of the independent variable *technique*, the discovery of selected objects was facilitated through *cut-away* or *fade-away* views, while the *baseline* condition required the observer to assume an occlusion-free viewpoint, i.e., close to the presenter.



(a) Experimental setup

(b) Virtual test environment

Figure 11.7: *The experimental setup for the study consisted of a two-user 3D display (about 3 meters wide) showing the engine compartment of a car at original scale. The experimental task involved the identification and finding of objects in this dense environment. Figures reprinted from [8] © 2010 IEEE*

Performance measures were captured in terms of the times needed for object discovery (during the presentation phase) and object retrieval. We also recorded user positions during the experiment to analyze the overall amount of movement and proxemic behavior. Between-subject comparisons with 8 participants per *technique* revealed significantly less movement and higher average interpersonal distances in both *show-through* conditions (Figure 11.8), while no significant differences in performance measures could be obtained [8].

In all conditions users maintained interpersonal distances between 76 and 120 cm (far phase of personal distances) for at least 50% of the time. Hall argued that people in professional collaboration prefer to stay further away from each other, namely in the close phase of social distance between 120 to 210 cm [125]. This is almost impossible if two users are moving in front of a 3 meter wide display. However, our *show-through* techniques made it possible to maintain such distances for almost the other 50% of time. The minimized participant movement during the presentation phase can be considered an indicator of emerging territoriality (see Section 6.3). In the *baseline* condition, instead, the participants needed to follow the experimenter closely, which resulted in more than twice as much overall movement and much smaller interpersonal distances. They could not even avoid intrusions to intimate distance (5.26% of time less than 46 cm distance).

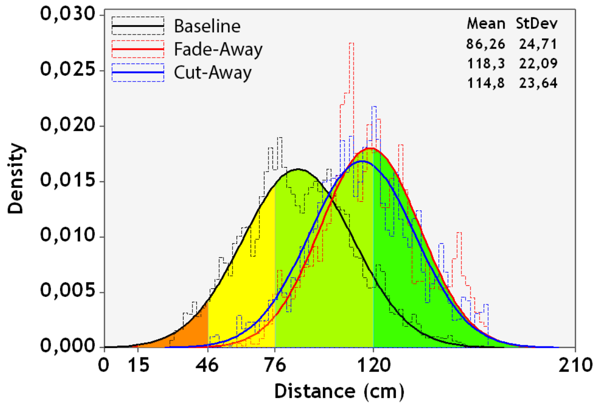


Figure 11.8: *The distribution of interpersonal distances per viewing technique. Dashed outlines represent the normalized histogram of logged distances. Solid lines show fitted normal distributions and shaded areas represent Hall's classes of interpersonal distances [125] (orange: private distance, yellow: close personal distance, light green: far personal distance, dark green: close social distance). Figure reprinted from [8] © 2010 IEEE*

11.2.2 Separate Views and Shared Perspectives

Many processes, and in particular creative ones, involve alternating phases of divergence and convergence. This is no different in cooperative action. Participants frequently switch between tight and loose coupling and thereby establish different territories for private activities and group exchange. The separate interaction with different objects requires sufficient space for concurrent manipulation without interference. If it comes to the individual exploration of a larger environment or the individual examination of the same object by multiple people, more elaborate means of coupling and decoupling become necessary.

Consider a joint tour through a historic architecture, a museum, or a botanic garden. Group members generally stay in sight of each other but they do not always move on the same path. Instead, each person pursues slightly different interests and explores other aspects of the environment. They also come together every now and then, to appreciate each other's discoveries. In collaborative virtual environments, however, independent navigation is not directly available. The tracked area for physical walking is generally not large enough. Individual virtual navigation, on the other hand, breaks the coherence of the shared workspace which results in perceptual conflicts (e.g. [60]).

We contemplated that separate interaction territories can satisfy the demand for temporary divergence of group activities. These can be realized as separate physical displays or virtual viewports. The window metaphor of 2D graphical user interfaces, for example, provides such independent viewports for parallel tasks. Virtual windows can also be used above each other as lenses that modulate visualization styles and interactive behavior [38] or to specify private interaction territories [340]. In the context of 3D applications, viewing windows and lenses have been suggested as portals that provide additional perspectives [277] and immediate access to distant virtual locations [327].

We implemented physical and virtual manifestations of such explicitly separated interaction territories for our collaborative virtual reality system and explored their use for conflict-free individual viewpoint navigation (Figure 11.9). The hardware solution is a multi-display infrastructure consisting of a multi-user 3D tabletop with multitouch input capabilities in vicinity of the multi-user 3D powerwall described in Chapter 10. The virtual version is a handheld portal window that follows the metaphor of virtual photography [189].

Figure 11.9: *A large 3D powerwall (back) and a multitouch 3D tabletop (front) serve as independent multi-user 3D viewports into a shared virtual world. Our framework combines both displays in a coherent 3D interaction space. A virtual 3D display, or portal (center, with white frame, see [189]) offers additional perspectives. The physical and virtual viewports serve for private interaction and group exchange. Their combination in a coherent workspace supports fluent transitions between tightly and loosely coupled cooperation. Here, a multi-scale 3D scan of prehistoric rock art and its environment (Valcamonica, Italy) is explored. The powerwall shows the rock surface at original scale. The tabletop user has adjusted a top-down view of the same rock panel at miniature scale. The user in the center used the handheld portal to create a scaled-up section view of an individual artwork (left) and shows it to the tabletop user (right).*

Both physical displays, the powerwall and the tabletop, provide three users with individual 3D stereoscopic views of a shared 3D scene. The underlying display technology could support up to six users (see chapter 10), but the tabletop is currently limited to three. The displays have been installed next to each other and they are running in synch to support the same group of users with two independent viewports. The 3D powerwall is equipped with the Spheron group navigation device (see Section 10.5). The tabletop view can be adjusted with multitouch 3D navigation techniques. On each display, all involved users perceive the same location of the shared virtual environment from their individual perspectives. If their interests diverge temporarily, they can use both devices separately and explore different regions. If they want to exchange information about their individual discoveries, they meet at one of both displays. The independent viewports serve as separate interaction territories that support rapid transitioning between tightly and loosely coupled cooperation.

The described setup is very well suited to simultaneously provide views of the same environment at different scales, e.g., overview and detail (Figure 11.9), or to support direct comparisons between two remote locations. Moving from one display to another, consequently, involves an immediate change of context. However, people may also want to deviate directly from a shared path or location, e.g., remain behind or continue in another direction. We developed a novel collaborative interaction technique that supports such behavior in multi-user virtual reality to a certain extent.

The Photoportals interaction technique allows to capture a current view of a shared location that can be stored for later use or it can be immediately adapted to facilitate the individual exploration of alternative perspectives at the scene with the current location as a starting point. The captured view is decoupled from the group navigation and thus supports individual exploration. Our interface design also facilitates the sharing of these perspectives and group navigation to the shown locations [189].

During a joint tour, users can capture interesting aspects of the environment using a dedicated tangible controller (Figure 11.10 a). The virtual photos are in fact 3D portals to the captured locations and support further adaptations of the viewport or virtual camera. The viewport can be freely manipulated, including 3D position, orientation, and scale, which enables perspectives that cannot be assumed through natural movement in front of the shared 3D display. In this way, an individual user can explore sideways, while the group is moving on. When that user reveals an interesting aspect, it can be easily shown to others (Figure 11.10 b).

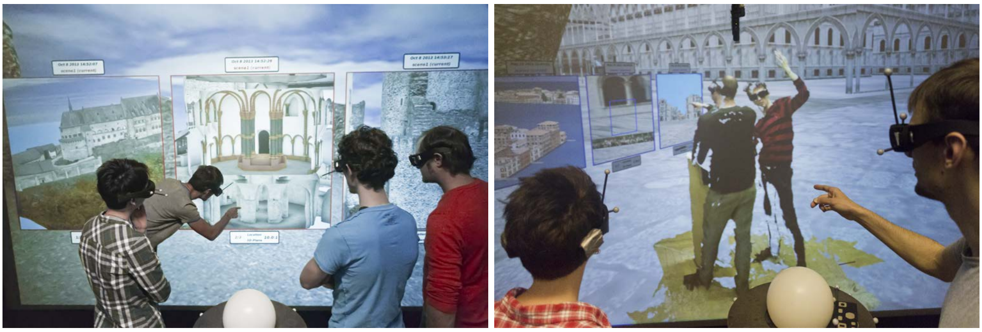
Photoportals also facilitate group navigation to captured locations. If everyone agrees, the size of the Photoportal can be increased until the visible scene on a physical display becomes replaced by the view of the Photoportal. This allows groups to travel effortlessly between locations in the virtual environment. We implemented a public gallery to facilitate the exchange about available Photoportals and agree on a new target location to enter (Figure 11.11). The public gallery is invoked and linked to group navigation by placing the tangible Photoportal controller on a tray of the

Spheron group navigation device. Users can also take a captured perspectives from one physical display and apply it to another, e.g., to explore the same location with a subgroup using different means.

(a) Capturing a virtual photo

(b) Sharing perspectives

Figure 11.10: *Photoportals enable the capturing of any 3D perspective in virtual environments (a). The position, orientation, and scale of the virtual camera remain adjustable, which allows the preparation of novel views to be shared with others (b). Figures reprinted from [189] © 2014 ACM; The 3D-model of castle Vianden is courtesy of ArcTron 3D GmbH (http://www.arctron.com)*



(a) A Photoportal gallery

(b) Remote users entering a Photoportal

Figure 11.11: *Following the metaphor of virtual photography, Photoportals offer a gallery of previously captured views. In combination with the Spheron group navigation device, this gallery can be shown to all participants and the group can decide to enter the shown location, by increasing the size of the selected portal. Figure (a) reprinted from [189] © 2014 ACM*

In the last section, we discussed issues arising from the fact that several features of a 3D scene are not equally visible to all users. Their joint exploration may require that both users assume very similar viewpoints, i.e. that they stay very close to each other, which can be undesirable. Our *show-through* techniques alleviate this issue by rendering parts of the occluding geometry transparent. This approach effectively supports the visual perception of hidden objects and their location from any perspective. Sometimes, however, the perspective matters, e.g., if the alignment of objects or the appearance of an object from a specific angle, is discussed. In these cases, Photoportals offer copies of a suitable view or viewport, such that others can perceive the scene from the same or a similar perspective (Figure 11.10 b). Moreover, Photoportals can create section views, since they clip all geometry in front of the virtual display (Figure 11.12 a). Also their shape can be adjusted, e.g. to a rectangular box exposing volumetric sections of the virtual environment (Figure 11.12 b). This allows users to uncover hidden objects with their geometric context and show them to others. Also user activities at the displayed locations can be observed and recorded (Figure 11.12 b). For many situations, Photoportals thus offer a viable alternative to the above described *show-through* techniques.

(a) A Photoportal as a cutting plane (b) A Photoportal as a section box

Figure 11.12: *Photoportals are dynamic. They support dynamic section views as well as the observation and capturing of user activities at any location in the shared virtual environment. The photo on the left shows the use of a Photoportal as a cutting plane (a). In the photo on the right a Photoportal box can be seen that shows a scaled section view to the activities of a remote user group in another area of the virtual castle. Figures reprinted from [189] © 2014 ACM; The 3D-model of castle Vianden is courtesy of ArcTron 3D GmbH (http://www.arctron.com)*

Multiple physical displays and virtual viewports can support territoriality in collaborative workspaces. They provide explicit interaction areas that can be used by individuals or subgroups to deviate from tightly coupled group activities. Group members can accomplish independent subtasks in parallel and contribute their results to the group's joint action. However, if the discoveries along different exploration paths cannot be effectively shared with others, such deviation can be ineffective and potentially impede further collaboration. Separate interaction spaces thus need to be tightly integrated with the shared workspace and support the spontaneous character of emergent territoriality.

The two physical 3D displays in our setup are both directly accessible to all users. They provide individual 3D views and sufficient interaction space. Photoportals also support multi-user 3D viewing, but their limited size and their handheld nature render them more suitable for private interaction. We observed that the exchange of views via Photoportals, in the sense of capturing perspectives (Figure 11.10 a) and entering the displayed locations (Figure 11.11), was key for fluent group interaction in our setup. The functionalities of physical and virtual viewports seem to complement each other. Our observations suggest that the combination of loosely coupled physical displays and highly dynamic virtual viewports increases the fluency and efficiency of collaborative 3D data analysis. A formal evaluation is pending. So far, we have been concerned with the usability of individual functionalities (see [189]). In the following we discuss further examples of complementary interaction capabilities and benefits of their combination.

11.3 Complementary Tools for Cooperative 3D Interaction

Complementarity is perhaps the one of most obvious foundations of successful cooperation. Instead of following each other blindly and doing the same, multiple actors with different views and capabilities can improve each other's awareness, comprehension, and action capabilities. This is not a given; it depends on the involved people and the situation at hand. However, people's inherently different skills and viewpoints favor social interaction with complementary roles. Complementary interface capabilities can facilitate and encourage such cooperative behavior.

We have argued earlier that *workspace coherence* and *emergent territoriality* enable groups to better exploit their members' individual contributions. Our experiments on collaborative search and memorization of environmental features, for example (see Chapter 10), demonstrated that groups can benefit considerably from individual views and direct gestural communication. Both can be provided through various interface configurations, e.g., a single shared display that accommodates all participants. We also observed that separate interaction areas, with extended means for virtual navigation, can increase the benefits of territoriality. Moreover, separate interaction spaces with complementary characteristics can positively affect the dynamics of territoriality. The suitability of different interface elements for varying activities is a reasonable motivation for frequent transitions between them.

Different display characteristic, as discussed above, afford different collaborative usage behavior. The tabletop in our setup facilitates direct face-to-face communication with application content located between equally involved users. 3D navigation with multitouch input, resembles object manipulation and does not seem to trigger cybersickness. The tabletop is thus particularly useful get an overview over a larger virtual environment and move rapidly between multiple sites. Our large vertical display with the group navigation device Spheron [187], on the other hand, affords immersive presentation settings with one user controlling the group's movement through the virtual environment. The handheld Photoportals seem to be most useful for private explorations and the exchange of individual discoveries. Users can choose the most suitable tools to achieve varying intermediate goals.

The simultaneous availability of different interaction facilities can encourage users to assume complementary roles. Instead of running into conflicts about control over the same functionalities or items, they can pursue different subtasks. This requires, that their individual contributions can be smoothly interleaved with common group activities or even combined to realize extended functionality. The above described integration of Photoportals in a multi-display environment, for example, simplifies group navigation between distant locations through the combination of complementary navigation interfaces. In the following section, we present two examples of complementary manipulation interfaces. One combines Photoportals with a pointing device, the other one builds on pointing devices with different functionalities.

11.3.1 Spatial References for Manipulation

Section 11.2.2 introduced Photoportals as separate viewports and navigation aids. Similar to Voodoo Dolls [267] or a world in miniature (WIM) [325], they can also provide users with an additional miniature representation of selected objects or fractions of the scene. Such secondary scene representations facilitate the visual examination of selected features. In this sense they offer an alternative to common object manipulation techniques like a pick ray or a virtual hand (see [43] for examples). In contrast to direct object manipulation in the shared 3D environment, the examination of a secondary representation, does not alter the shared view at a scene or the scene itself and is therefore less disturbing to others.

Photoportals and direct manipulation tools offer complementary functionalities and their combination enables several benefits. For example, direct manipulation is most effective if the object is located within arm's reach and fits comfortably between both hands. Photoportals can serve as adequate spatial references for the manipulation of objects that would be otherwise too large, too small, or too far away.

The basic principle has been described previously by Stoev and Schmalstieg [327]. When a user is operating through the portal, manipulation input is transformed and applied at the represented location. This facilitates adjustments to the placement of distant objects (Figure 11.13 b). They can also be dragged across the boundaries of the virtual viewport for a direct transfer between the represented location and the main scene (Figure 11.13 a). Our implementation of Photoportals for collaborative virtual reality extended and refined the work of Stoev and Schmalstieg. Most importantly, we developed an elaborate user interface for the creation, management, and exchange of secondary scene representations [189]. The operation of Photoportals and further manipulation tools like the virtual pick ray can be easily distributed between two hands or, for more complex actions, between two users (Figure 11.13).

(a) Drag and drop between remote locations (b) Remote object manipulation

Figure 11.13: *Photoportals provide access to remote objects. In left side photo (a) a ship model is extracted from a Photoportal presented by another user. The Photo on the right side (b) shows two local (left) and two remote users (right) in a harbor scene. One of the remote users manipulates the placement of the ship in the background using a box-shaped Photoportal, which in turn is operated by a local user and currently coupled to the Spheron. Figure reprinted from [189] © 2014 ACM*

Photoportals can also be combined with the Spheron group navigation device. Placing the tangible camera controller on its tray switches from private to public use. In planar mode this activates the public gallery and enables the whole group to chose a target destination (Figure 11.11). In box mode, the volumetric section view will be placed above the large trackball (Figure 11.13 b). Rotation input to the trackball is then directly applied to control the orientation of the Photoportal box. This can be useful to minimize the physical effort of holding the device and, more importantly, the fixed location facilitates examination by more than one user.

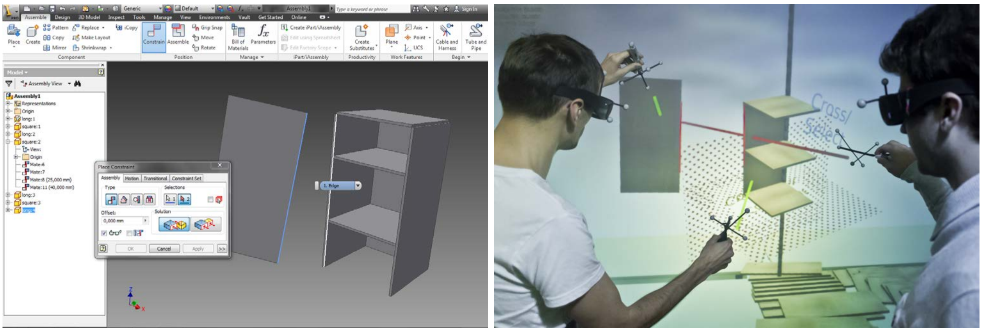
We have not evaluated so far whether the described combinations of Photoportals with other interaction techniques like ray-based manipulation and group navigation enable performance benefits in any particular task. Our observations during experimental test sessions, showed that the resulting interaction possibilities were frequently used and appreciated (e.g. [189]). During the European research project 3D-Pitoti, we implemented techniques for the collaborative analysis of 3D scanning data from prehistoric rock art and the surrounding landscape [1]. In the developed application prototypes Photoportals were used to facilitate the comparison of features in various scanned rock surfaces (Figure 11.9). To this end, they could be used in combination with a virtual torch that improved the visibility of shallow 3D structures through highlights and shadows. Archaeologists testing our demonstrator in day-long workshops agreed that both tools and their combination were key elements of the system.

11.3.2 Complementary Tools for Cooperative 3D Object Assembly

We believe that our novel technologies for multi-user 3D visualization and collaborative interaction can be useful for several productivity applications. As a proof of concept, we implemented a basic set of complementary tools for the cooperative assembly of 3D objects in computer aided design (CAD). Desktop-based CAD applications (e.g., [14, 73]) often include tools for the assembly of 3D objects based on geometric constraints (Figure 11.14 a). The process generally requires the rough 3D placement of objects and the specification of relations between their geometric features. Manual 3D placement alone is not accurate enough and individual geometric constraints do not fully specify the relation of two objects. The constraint solver applies a solution that is closest to their initial relation, hence they should be roughly placed as desired. Moreover, 3D view navigation or object manipulation is necessary to see and access the relevant geometrical features. In desktop applications, only a mouse is available to select and move 3D objects, hence all steps must be performed sequentially. The keyboard can be used for mode switching, e.g., between pointer operation and view navigation as described in Chapter 8. The alignment of two objects with these tools requires at least the following steps:

1. Object translation to the approximate target location
2. Object rotation to the approximate target orientation
3. View navigation
4. Feature selection
5. Constraint selection from a menu
6. Invocation of the constraint solver

The object manipulation steps (1 & 2) typically involve repetitions, because the 2D interface cannot operate the three degrees of freedom simultaneously. In addition, between the different steps the view must generally be changed more frequently to gain access to the relevant objects and geometric features. 3D interaction techniques in immersive environments, instead, enable object movement with simultaneous control of 3D rotation and translation. Even multiple geometries can be moved at the same time, e.g., to roughly align two objects at a comfortable location. Therefore, step one and two can potentially be performed in a single action. Moreover, users can simply walk around the displayed objects to assume or maintain an optimal view during their actions. Last, but not least, cooperating users can distribute the subtasks of object manipulation and constraint specification between each other (Figure 11.14 b). This parallelization of complementary subtasks suggests that cooperative 3D object assembly in immersive virtual reality can be considerably more effective than its desktop-based counterpart.



(a) Autodesk Inventor [14]

(b) Cooperative 3D object assembly

Figure 11.14: *Our experimental application for cooperative 3D object assembly shown on the right (b) offers a subset of the tools for constraint-based assembly design in desktop based CAD applications like Autodesk Inventor [14] (a). Multi-user virtual reality enables the simultaneous 3D manipulation of multiple objects and the distribution of subtaks between participants. Here, the user on the left controls the placement of two objects, while the user on the right defines geometric constraints.*

We implemented a very basic test application for cooperative 3D assembly design with three complementary interface functionalities. Two users could assume individual viewpoints at the shared 3D scene by walking around. Two 3D pointing devices were used for ray-based 3D object manipulation. Another 3D pointing device offered the complementary functionality to specify constraints. This subtask required the initial selection of a constraint type from a menu (e.g., adjacency, parallelism, orthogonality, etc.) and the subsequent selection of the faces or edges between which the rule should be applied. Object faces were selected through intersections with the ray pointer, while edges had to be crossed with the 3D ray pointer for selection. When two objects with specified constraints were simultaneously moved towards each other and released in spatial proximity, they were automatically connected according to the specified constraint. The constraint solver was based on the bullet physics engine [64].

We experimentally used the system to assemble simple furniture models from a given set of virtual boards (Figure 11.14). One user was generally operating the two manipulation devices, while another defined the constraints. The users coordinated their actions primarily using words and pointing gestures. The subtask of connecting two boards generally consisted of three steps. The manipulating user first took two virtual boards and oriented them towards the other one, such that the desired constraint could be applied most comfortably. When the constraint was defined, the manipulating user roughly connected the two parts and then released them for automatic alignment. Preliminary results indicate that this process takes less than half the time as required for one user to achieve the same results with a professional desktop-based CAD application.

Chapter 12

Conclusions and Future Work

This Ph.D. thesis explored the foundations, the design, and the evaluation of user interfaces for cooperation. Synergistic cooperation can be consistently observed in interactions with several degrees of freedom. Chapter 4 showed its fundamental role in human movement coordination. Chapter 5 compiled representative examples of multimodal, bimanual, and collaborative user interfaces that build on cooperative interaction. Following this common thread of prior research and developments, this thesis suggested to emphasize the paradigm of cooperation as an essential building block for sophisticated human-computer interfaces and interaction techniques.

Chapter 6 illustrated the potential benefits with common cooperation patterns. *Equivalence* highlights the availability of multiple interaction opportunities with at least comparable expressive capabilities. This in turn offers higher situational flexibility including the ad-hoc collaboration of user groups through multiple access points. The combination of *equivalent* actions and events supports mutual *confirmation and intensification*. The *specialization* of tools, skills, and interaction roles, or the *transfer* of intermediate results, on the other hand, facilitate more expressive interaction through the *complementarity* of multiple contributions.

Until recently, the design of most computer workplaces has focused on individual users and often with only few concurrent input options. When interacting with a desktop computer or with mobile devices, users tend to turn away from their collocated peers. The limited size alone of personal computer interfaces does not accommodate multiple users. Larger displays are commonly used for public presentations and advertising, however, these rarely support interactivity beyond showing and watching. Fluent transitioning between private interaction and group exchange is generally not supported. Current developments in the fields of ubiquitous computing and collaborative work promise the interactive coupling of personal and public devices that may solve the latter issue (e.g. [71, 85, 86, 110, 227, 276, 365]). However, interface and interaction design paid only little attention, so far, on effective support for cooperative action. Above all, current interfaces lack meaningful interpretations of concurrent user input.

Chapter 2 discussed the performance limitations of single-handed motion input, which is the default input paradigm for most computer applications. In particular, a comparison of different studies on user performance in aimed movement tasks indicated that effective 3D motion input with computer interfaces seems to be difficult to achieve. 3D motion control involves a higher number of degrees of freedom than its 2D counterpart. Chapter 3, therefore reviewed research on the perceptual integrality and separability of geometrical dimensions and their relation to the control of multiple degrees of freedom of movement. The review indicated inconsistent interpretations of Garner's theory on perceptual integrality and separability [102] in HCI research. Its applicability to active motion control seems to be limited. Research on perception-action coupling, and stimulus-response compatibility in particular, enables a better understanding of the cognitive processes involved in motion input to graphical user interfaces.

This thesis argues that cooperative user interfaces can enable in more effective and more expressive human-computer interaction, but, effective cooperation is a challenge in itself. The coordination overhead can be particularly demanding if multiple people are involved. To gain a better understanding of the opportunities and challenges of cooperative interaction, Chapter 4 reviewed research on human motor control and interpersonal coordination. Chapter 5 discussed the benefits and drawbacks of prior user interfaces that build on cooperative action. Chapter 6 considered the applicability of generic cooperation patterns and derived the most essential requirements for cooperative action. This theoretical work led to the development of three high-level design principles for cooperative user interfaces: *workspace coherence*, *complementary capabilities*, and *emergent territoriality*.

The design and evaluation of cooperative user interfaces can be approached from various angles. Part II of this thesis described three very different examples: the analysis of detailed effects of devices and transfer functions used in bimanual interface design (Chapter 8), the design and evaluation of a novel multi-touch input gesture based on knowledge on bimanual movement coordination (Chapter 9), and the development and evaluation of a novel 3D display technology for multiple collocated users (Chapter 10).

Chapter 8 explored detailed effects of interface design on the speed and accuracy of bimanual cooperation in desktop-based 3D applications. The results demonstrated higher input accuracy for interface designs that supported a more balanced bimanual workload. We also observed effects of input device characteristics and transfer functions on the course of bimanual interactions. With an elastic rate controller in one hand, users tended to maintain a continuous movement while aiming at moving targets with the other. Cooperative bimanual input with isotonic devices for position control in both hands, instead, was more sequential and supported even higher accuracy.

Chapter 9 reported the analysis of temporal patterns during bimanual interaction with multitouch devices. The findings, in correspondence to prior research, suggested that symmetric and asymmetric input can be distinguished based on their temporal differences of action onset of both input streams. We applied these results to the design of a novel input gesture that fosters *complementarity* through the *specialization* of input from both hands. Commonly, touch input from one or more fingers is applied to either view navigation or to object selection and manipulation. Dwell times are used for mode switching. Our novel approach allows the concurrency of both types of input and demonstrated considerable performance benefits in comparison to the status quo.

Chapter 10 described the development and evaluation of a novel 3D display technology that provides individual 3D views for up to six users. We used the technology to realize a coherent mixed-reality workspace for multi-user collaboration. Up to six collocated users can experience a shared 3D environment and their mutual presence in this context. They can directly interact with each other and the displayed content since they are provided with individual, perspectively correct views. A formal evaluation of the system in a collaborative visual search task proved that multiple users could take advantage of their possibility to cooperate. We also revealed design challenges that result from the application of this novel technology and suggested solutions for different use cases.

Finally, Part III of this thesis showed the relevance and applicability of our design principles, *workspace coherence*, *complementary capabilities*, and *emergent territoriality* (see Chapter 6), with examples of cooperative user interfaces for multi-user virtual reality. All these examples were realized on the basis of multi-user 3D display technology as described in Chapter 10. The experimental systems facilitated the creation of a coherent interaction space for multiple collocated users.

Section 11.1 showed how *workspace coherence* can also be achieved in telepresence settings. Our observations emphasized the value and importance of meeting in a shared 3D environment to enable the meaningful exchange with others about the application content. Our experimental results demonstrated the immediate comprehensibility of gestural communication and body language in such a setting.

Section 11.2 demonstrated examples of interface support for *emergent territoriality* to conform with behavioral norms and to enable temporary deviations from group activities. One experimental system showed how crowding around a viewpoint towards objects of interest can be avoided in virtual reality through personalized visual presentations of the shared interaction space. Another series of experiments explored the tight integration of multiple independent viewports in a coherent interaction space. The resulting system supported *emergent territoriality*, i.e., the spontaneous establishment and dynamic management of dedicated spaces for private interaction, storage, and group exchange, which in turn facilitated fluent transitions between loosely and tightly coupled cooperation.

Section 11.3 presented two examples of mutual support through *complementary capabilities*. More specifically, it was suggested that users can support each other through the provision of spatial references for the comparison of views, the exchange of objects, and effortless 3D navigation. Another case study presented an experimental system for collaborative 3D assembly design, which allowed two users to perform a virtual object assembly task more effectively together. A suite of cooperative 3D manipulation tools supported their cooperation with the complementary roles of direct object manipulation and the specification of geometric constraints.

12.1 Thesis Contributions

Above all, this thesis contributes a new perspective on interaction design with a focus on synergistic cooperation. It emphasized the pivotal role of cooperative action in several established user interfaces and prior research prototypes (see Chapter 5) and made a first attempt to define cooperation as a universal design paradigm for a broad range of interactive setups: from multimodal and bimanual user interfaces to the collaboration of multiple users.

The discussion of related research in psychology and human-computer interaction contributes a compilation of the cognitive, behavioral, and ergonomic foundations of cooperative user interfaces. Moreover, we have discussed the applicability of generic cooperation patterns and derived high-level design principles for cooperative user interfaces. The identification of *workspace coherence*, *emergent territoriality*, and *complementary capabilities* as crucial aspects of cooperative user interfaces is another major contribution. A series of case studies demonstrated their relevance and applicability in the context of multi-user virtual reality.

On a more concrete level, the work in this thesis involved the design, development, and evaluation of three interface prototypes which were reported in Part II and additional case studies that were described in Part III. Each of these developments constitutes an individual contribution to the field. Together, they support the working hypotheses H1-H3 stated in Chapter 7.

H1 stated that “interface support for cooperative action is beneficial for single users and, at the same time, facilitates the collaboration of groups.” The review of related work had already identified several supporting examples of existing user interfaces and prior research prototypes in the fields of multimodal interaction, bimanual interfaces and multi-user collaboration. The reported investigations in the fields of desktop-based 3D applications, mobile multitouch devices, and collaborative virtual reality further substantiate the claim. The experimental results described in Chapter 8 and 9 demonstrated significant benefits of cooperative bimanual interaction. The interface developments for multi-user virtual reality (Chapter 10 & 11) applied highly similar cooperation patterns to the collaboration of multiple users. Consequently, collaborative 3D user interfaces like the Spheron [187], Photoportals [189], and their combinations with ray-based manipulation input [189], can be operated by a single user with both hands or the subtasks can be distributed between different users.

H2 stated that "the combination of concurrent user actions enables improved workload balancing, higher interaction fluency, more flexibility, and extended interface functionalities." The reported research and developments support this hypothesis. The experiments on bimanual input cooperation with desktop-based 3D devices (Chapter 8) revealed significantly higher input accuracy and more robust performance across different subtasks. The novel multitouch input technique Hold-and-Move (Chapter 9) enabled more fluent interaction through extended input options. The multi-user 3D display system (Chapter 10) enabled new possibilities of cooperative information exchange between multiple users through direct gestural communication, which could be proven with significantly better performance in a visual search task. The system constitutes a step change for collaborative virtual reality. It extends the functionality of existing stereo displays to accommodate up to six independent viewers.

H3 stated that "The design of cooperative user interfaces can be successfully guided by the high-level design principles *workspace coherence*, *emergent territoriality*, and *complementary capabilities*." The work described in Part II supports individual aspects of these principles. Both experiments on bimanual interaction (Chapter 8 & 9), for example, revealed significantly improved task performance based on *complementary capabilities*. The novel multi-user 3D display technology (Chapter 10) enables *workspace coherence* and *emergent territoriality* for collaborative virtual reality in collocated settings. A number of case studies with this display technology, described in Part III, confirmed the expected advantages. Moreover, both systems for cooperative object manipulation (Section 11.3) emphasized the benefits of *complementary interface capabilities*.

12.2 Follow-Up Research

This thesis explored the design of user interfaces with a focus on support for cooperative action. The presented research answered some essential questions and uncovered others that imply new directions for follow-up research and development.

The design principles of cooperative user interfaces are currently only broadly defined. Real-world interaction in collocated settings is considered the gold standard (e.g. [149,187,311]), that should be achieved and that can be further improved through interactive technologies (e.g. [8, 189]). However, several applications do not enable real-world interaction as a baseline. Telepresence setups, for example [26], imply spatiotemporal inconsistencies. The concept of *workspace coherence* incorporates flexibility in that regard, but it requires further research to isolate the most relevant parameters and acceptable tolerances for different types of applications. Gutwin and Greenberg's framework on workspace awareness [121] offers a starting point in that direction. In terms of territorial behavior and proxemics, we need to consider that telepresence applications imply the superimposition of different physical places, each with multiple collocated people that may be more or less involved and highly varying physical affordances.

The review of aimed movement studies in Chapter 2 indicated that user performance may depend heavily on the number of involved degrees of freedom, the type of transformation, and the availability of physical support. Fitts's Law, instead, considers only the target distance and the target width as external factors [87]. We are currently preparing studies on this topic. In this regard, we will also explore synergies of cooperation in aimed movements. Research on motor control suggests mutual error compensation, in particular, during complex movements with multiple degrees of freedom (see Chapter 4).

This thesis emphasized the relevance of *complementary interface capabilities* to support cooperative action. However, the interpretation of concurrent input actions and their combinations can be challenging. The number of interface states grows exponentially with the number of distinct inputs that can be combined. We are working on design patterns that facilitate this state management, e.g., through the exchange of higher-level interaction requests and their resolution in the application objects.

In this context, we are currently implementing variations of the Hold-and-Move interaction pattern (Chapter 9) to facilitate hierarchical 3D object manipulation. Furthermore, we are planning to explore its applicability to distinguish separate inputs from multiple users, e.g., during collaboration with a multitouch-enabled tabletop display.

So far, our work has focused on cooperative interaction with a single, possibly distributed, application. However, users bring their own devices and applications to meetings with others. The integration of personal devices as further access points for interaction and data exchange promises increased fluency of cooperative work. Moreover, the content of collaborative applications becomes more relevant, if the participants can access it at any time with their personal devices and share it with further people.

We believe that the paradigm of cooperative user interfaces can improve the effectiveness, ergonomics, and social compatibility of computer workplaces. These potential benefits can be best achieved through the involvement of users in the design processes. This in turn requires the development of application prototypes that are meaningful to users. We will continue our research on cooperative user interfaces in the domains of visual data analysis, learning, decision making, telecommunication, and creative productivity.

Bibliography

[1] 3D-Pitoti. 3d acquisition, processing and presentation of prehistoric european rock-art – ec fp7 project (ict-2011-600545). `http://www.3d-pitoti.eu/`, 2013-2016 (accessed May 8, 2016).

[2] Accot, J., and Zhai, S. Beyond fitts' law: Models for trajectory-based hci tasks. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1997), CHI '97, ACM, pp. 295–302.

[3] Accot, J., and Zhai, S. Refining fitts' law models for bivariate pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2003), CHI '03, ACM, pp. 193–200.

[4] Agrawala, M., Beers, A. C., McDowall, I., Fröhlich, B., Bolas, M., and Hanrahan, P. The two-user responsive workbench: Support for collaboration through individual views of a shared space. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1997), SIGGRAPH '97, ACM Press/Addison-Wesley Publishing Co., pp. 327–332.

[5] Aguerreche, L., Duval, T., and Lécuyer, A. 3-handed manipulation of virtual objects. In *Proceedings of the 15th Joint Virtual Reality Eurographics Conference on Virtual Environments* (Aire-la-Ville, Switzerland, Switzerland, 2009), JVRC'09, Eurographics Association, pp. 153–156.

[6] Aguerreche, L., Duval, T., and Lécuyer, A. Evaluation of a Reconfigurable Tangible Device for Collaborative Manipulation of Objects in Virtual Reality. In *Theory and Practice of Computer Graphics* (Warwick, United Kingdom, Sept. 2011), U. E. Chapter, Ed., pp. 81–88.

[7] Andersen, T. H. A simple movement time model for scrolling. In *CHI '05: CHI '05 extended abstracts on Human factors in computing systems* (New York, NY, USA, 2005), ACM, pp. 1180–1183.

[8] Argelaguet, F., Kunert, A., Kulik, A., and Froehlich, B. Improving co-located collaboration with show-through techniques. In *3D User Interfaces (3DUI), 2010 IEEE Symposium on* (march 2010), pp. 55 –62.

[9] ARSENAULT, R., AND WARE, C. The importance of stereo and eye-coupled perspective for eye-hand coordination in fish tank vr. *Presence: Teleoper. Virtual Environ. 13*, 5 (Oct. 2004), 549–559.

[10] ARTHUR, K., PRESTON, T., TAYLOR, R., BROOKS, F., WHITTON, M., AND WRIGHT, W. Designing and building the pit: a head-tracked stereo workspace for two users. In *Proceedings of IPT '98, Workshop CD-ROM* (1998).

[11] ASHBY, F. G., AND MADDOX, W. T. A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology 38*, 4 (1994), 423–466.

[12] ASPIN, R., AND ROBERTS, D. A model for distributed, co-located interaction in urban design/review visualisation. In *In Proceedings of IEEE DS-RT 2005* (oct. 2005), pp. 70–77.

[13] AU, O. K.-C., AND TAI, C.-L. Multitouch finger registration and its applications. In *Proceedings of the 22Nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (New York, NY, USA, 2010), OZCHI '10, ACM, pp. 41–48.

[14] AUTODESK. Inventor. `http://www.autodesk.de/products/inventor`, 2016 (accessed May 8, 2016).

[15] AVANGONG. Cross-platform software framework for virtual reality. `https://github.com/vrsys/avango`, 2016 (accessed March 5, 2016).

[16] AZAD, A., RUIZ, J., VOGEL, D., HANCOCK, M., AND LANK, E. Territoriality and behaviour on and around large vertical publicly-shared displays. In *Proceedings of the Designing Interactive Systems Conference* (New York, NY, USA, 2012), DIS '12, ACM, pp. 468–477.

[17] BACH, P., AND TIPPER, S. P. Implicit action encoding influences personal-trait judgments. *Cognition 102*, 2 (2007), 151 – 178.

[18] BAKER, H. H., BHATTI, N., TANGUAY, D., SOBEL, I., GELB, D., GOSS, M. E., CULBERTSON, W. B., AND MALZBENDER, T. Understanding performance in coliseum, an immersive videoconferencing system. *ACM Trans. Multimedia Comput. Commun. Appl. 1*, 2 (May 2005), 190–210.

[19] BAKER, K., GREENBERG, S., AND GUTWIN, C. Heuristic evaluation of groupware based on the mechanics of collaboration. In *Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction* (London, UK, UK, 2001), EHCI '01, Springer-Verlag, pp. 123–140.

[20] BAKER, K., GREENBERG, S., AND GUTWIN, C. Empirical development of a heuristic evaluation methodology for shared workspace groupware. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 2002), CSCW '02, ACM, pp. 96–105.

[21] BALAKRISHNAN, R., BAUDEL, T., KURTENBACH, G., AND FITZMAURICE, G. The rockin'mouse: Integral 3d manipulation on a plane. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1997), CHI '97, ACM, pp. 311–318.

[22] BALAKRISHNAN, R., AND HINCKLEY, K. The role of kinesthetic reference frames in two-handed input performance. In *UIST '99: Proceedings of the 12th annual ACM symposium on User interface software and technology* (New York, NY, USA, 1999), ACM, pp. 171–178.

[23] BALOGH, T., AND KOVÁCS, P. T. Real-time 3d light field transmission. In *Proc. of SPIE 2010* (2010), vol. 7724, pp. 772406–772406–7.

[24] BARCO. Virtual surgery table. `http://www.barco.com/en/virtualreality/product/443`, 1999 (accessed May 9, 2016).

[25] BAUDISCH, P., CUTRELL, E., HINCKLEY, K., AND EVERSOLE, A. Snap-and-go: Helping users align objects without the modality of traditional snapping. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2005), CHI '05, ACM, pp. 301–310.

[26] BECK, S., KUNERT, A., KULIK, A., AND FRÖHLICH, B. Immersive group-to-group telepresence. *Visualization and Computer Graphics, IEEE Transactions on 19*, 4 (2013), 616–625.

[27] BENFORD, S., BEDERSON, B. B., ÅKESSON, K.-P., BAYON, V., DRUIN, A., HANSSON, P., HOURCADE, J. P., INGRAM, R., NEALE, H., O'MALLEY, C., SIMSARIAN, K. T., STANTON, D., SUNDBLAD, Y., AND TAXÉN, G. Designing storytelling technologies to encouraging collaboration between young children. In *Proc. of CHI 2000* (2000), ACM Press, pp. 556–563.

[28] BENFORD, S., GREENHALGH, C., RODDEN, T., AND PYCOCK, J. Collaborative virtual environments. *Commun. ACM 44* (July 2001), 79–85.

[29] BENKO, H., AND FEINER, S. Balloon selection: A multi-finger technique for accurate low-fatigue 3d selection. In *2007 IEEE Symposium on 3D User Interfaces* (March 2007).

[30] BENKO, H., JOTA, R., AND WILSON, A. Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proc. of CHI 2012* (New York, NY, USA, 2012), ACM Press, pp. 199–208.

[31] BÉRARD, F., IP, J., BENOVOY, M., EL-SHIMY, D., BLUM, J. R., AND COOPERSTOCK, J. R. Did "minority report" get it wrong? superiority of the mouse over 3d input devices in a 3d placement task. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II* (Berlin, Heidelberg, Germany, 2009), INTERACT '09, Springer-Verlag, pp. 400–414.

[32] BERNSTEIN, N. [studies on biomechanics of hitting using optical recordings]. *Annals of the Central Institute of labor 1* (1923), 19–79.

[33] BERNSTEIN, N. A. *The Co-ordination and regulation of movements*, first english edition ed. Pergamon Press Ltd., 1967.

[34] BERTHOZ, A. *The brain's sense of movement: perspectives in cognitive neuroscience*. Harvard University Press, 2000.

[35] BIER, E. A. Skitters and jacks: Interactive 3d positioning tools. In *Proceedings of the 1986 Workshop on Interactive 3D Graphics* (New York, NY, USA, 1987), I3D '86, ACM, pp. 183–196.

[36] BIER, E. A. Snap-dragging in three dimensions. *SIGGRAPH Comput. Graph. 24*, 2 (Feb. 1990), 193–204.

[37] BIER, E. A., AND STONE, M. C. Snap-dragging. *SIGGRAPH Comput. Graph. 20*, 4 (Aug. 1986), 233–240.

[38] BIER, E. A., STONE, M. C., PIER, K., BUXTON, W., AND DEROSE, T. D. Toolglass and magic lenses: The see-through interface. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1993), SIGGRAPH '93, ACM, pp. 73–80.

[39] BIMBER, O., FROEHLICH, B., SCHMALSTIEG, D., AND ENCARNACAO, L. M. The virtual showcase. *IEEE Computer Graphics and Applications 21* (2001), 48–55.

[40] BOLAS, M., MCDOWALL, I., AND CORR, D. New research and explorations into multiuser immersive display systems. *IEEE Computer Graphics and Applications 24* (2004), 18–21.

[41] BOLT, R. A. "put-that-there": Voice and gesture at the graphics interface. In *SIGGRAPH '80: Proceedings of the 7th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1980), ACM, pp. 262–270.

[42] BOURDOT, P., CONVARD, T., PICON, F., AMMI, M., TOURAINE, D., AND VÉZIEN, J. M. Vr-cad integration: Multimodal immersive interaction and advanced haptic paradigms for implicit edition of cad models. *Comput. Aided Des. 42*, 5 (May 2010), 445–461.

[43] BOWMAN, D., KRUJIFF, E., VIOLA, J. L., AND POUPYREV, I. *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2005.

[44] BRANDL, P., FORLINES, C., WIGDOR, D., HALLER, M., AND SHEN, C. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *Proc. AVI 2008* (2008), ACM Press, pp. 154–161.

[45] BRASS, M., BEKKERING, H., AND PRINZ, W. Movement observation affects movement execution in a simple response task. *Acta psychologica 106*, 1 (2001), 3–22.

[46] BRICKER, L. J., BAKER, M. J., AND TANIMOTO, S. L. Support for cooperatively controlled objects in multimedia applications. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems* (1997), CHI EA '97, ACM Press, pp. 313–314.

[47] BROLL, W., MEIER, E., AND SCHARDT, T. The virtual round table - a collaborative augmented multi-user environment. In *Proc. of CVE 2000* (2000), CVE '00, ACM Press, pp. 39–45.

[48] BUXTON, W. Chunking and phrasing and the design of human-computer dialogues. In *Proceedings of the IFIP World Computer Congress* (1986), North Holland Publishers, pp. 475–480.

[49] BUXTON, W. A. S. Telepresence: integrating shared task and person spaces. In *Proc. of Graphics Interface 1992* (San Francisco, CA, USA, 1992), Morgan Kaufmann Publishers Inc., pp. 123–129.

[50] CAO, X., MASSIMI, M., AND BALAKRISHNAN, R. Flashlight jigsaw: An exploratory study of an ad-hoc multi-player game on public displays. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 2008), CSCW '08, ACM, pp. 77–86.

[51] CARD, S. K., MORAN, T. P., AND NEWELL, A. The keystroke-level model for user performance time with interactive systems. *Commun. ACM 23*, 7 (July 1980), 396–410.

[52] CARD, S. K., NEWELL, A., AND MORAN, T. P. *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1983.

[53] CARLSSON, C., AND HAGSAND, O. Dive a multi-user virtual reality system. In *Proceedings of the 1993 IEEE Virtual Reality Annual International Symposium* (Washington, DC, USA, 1993), VRAIS '93, IEEE Computer Society, pp. 394–400.

[54] CARSON, R. G., THOMAS, J., SUMMERS, J. J., AND WALTERS & ANDRAS SEMJEN, M. R. The dynamics of bimanual circle drawing. *The Quarterly Journal of Experimental Psychology: Section A 50*, 3 (1997), 664–683.

[55] CASIEZ, G., VOGEL, D., BALAKRISHNAN, R., AND COCKBURN, A. The Impact of Control-Display Gain on User Performance in Pointing Tasks. *Human-Computer Interaction 23*, 3 (2008), 215–250.

[56] CASIEZ, G., VOGEL, D., PAN, Q., AND CHAILLOU, C. Rubberedge: Reducing clutching by combining position and rate control with elastic feedback. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2007), UIST '07, ACM, pp. 129–138.

[57] CHA, Y., AND MYUNG, R. Extended fitts' law for 3d pointing tasks using 3d target arrangements. *International Journal of Industrial Ergonomics 43*, 4 (2013), 350 – 355.

[58] CHAPUIS, O., AND DRAGICEVIC, P. Effects of motor scale, visual scale, and quantization on small target acquisition difficulty. *ACM Trans. Comput.-Hum. Interact. 18*, 3 (Aug. 2011), 13:1–13:32.

[59] CHEN, M., MOUNTFORD, S. J., AND SELLEN, A. A study in interactive 3-d rotation using 2-d control devices. In *SIGGRAPH '88: Proceedings of the 15th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1988), ACM, pp. 121–129.

[60] CHEN, W., CLAVEL, C., FÉREY, N., AND BOURDOT, P. Perceptual conflicts in a multi-stereoscopic immersive virtual environment: Case study on face-to-face interaction through an avatar. *Presence: Teleoperators and Virtual Environments 23*, 4 (2014), 410–429.

[61] CLAUDE MARTIN, J. Tycoon: Theoretical framework and software tools for multimodal interfaces. In *In John Lee (Ed.), Intelligence and Multimodality in Multimedia Interfaces* (1998), AAAI Press.

[62] COHÉ, A., DÈCLE, F., AND HACHET, M. tbox: A 3d transformation widget designed for touch-screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2011), CHI '11, ACM, pp. 3005–3008.

[63] COSSAIRT, O. S., NAPOLI, J., HILL, S. L., DORVAL, R. K., AND FAVALORA, G. E. Occlusion-capable multiview volumetric three-dimensional display. *Appl. Opt. 46*, 8 (Mar 2007), 1244–1250.

[64] COUMANS, E. Bullet physics engine. `http://bulletphysics.org`, 2016 (accessed May 8, 2016).

[65] CRUZ-NEIRA, C., SANDIN, D. J., DEFANTI, T. A., KENYON, R. V., AND HART, J. C. The cave: audio visual experience automatic virtual environment. *Commun. ACM 35* (June 1992), 64–72.

[66] CSIKSZENTMIHALYI, M. The psychology of optimal experience new york, 1990.

[67] CUTLER, L. D., FRÖHLICH, B., AND HANRAHAN, P. Two-handed direct manipulation on the responsive workbench. In *Proc. I3D 1997* (1997), ACM Press, pp. 107–114.

[68] DAFFERTSHOFER, A., LAMOTH, C. J., MEIJER, O. G., AND BEEK, P. J. Pca in studying coordination and variability: a tutorial. *Clinical biomechanics 19*, 4 (2004), 415–428.

[69] DAMERA-VENKATA, N., AND CHANG, N. L. Realizing super-resolution with superimposed projection. In *CVPR: Computer Vision and Pattern Recognition* (2007).

[70] DANG, C., STRAUB, M., AND ANDRÉ, E. Hand distinction for multi-touch tabletop interaction. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (2009), ACM, pp. 101–108.

[71] DANG, C. T., AND ANDRÉ, E. A framework for the development of multi-display environment applications supporting interactive real-time portals. In *Proceedings of the 2014 ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (New York, NY, USA, 2014), EICS '14, ACM, pp. 45–54.

[72] D'ANGELO, D., WESCHE, G., FOURSA, M., AND BOGEN, M. The benefits of co-located collaboration and immersion on assembly modeling in virtual environments. In *Advances in Visual Computing*, vol. 5358 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2008, pp. 478–487.

[73] DASSAULT. Catia. `http://www.3ds.com/products-services/catia`, 2016 (accessed May 8, 2016).

[74] DAY, R. S., AND WOOD, C. C. Interactions between linguistic and nonlinguistic processing. *Journal of the Acoustical Society of America 51*, 79 (1972), 247–252.

[75] DJAJADININGRAT, T., MATTHEWS, B., AND STIENSTRA, M. Easy doesn't do it: Skill and expression in tangible aesthetics. *Personal Ubiquitous Comput. 11*, 8 (Dec. 2007), 657–676.

[76] DODGSON, N. A. Autostereoscopic 3d displays. *Computer 38*, 8 (2005), 31–36.

[77] DOMKIN, D., LACZKO, J., JARIC, S., JOHANSSON, H., AND LATASH, M. L. Structure of joint variability in bimanual pointing tasks. *Experimental brain research 143*, 1 (2002), 11–23.

[78] DREWES, H. Only one fitts' law formula please! In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2010), CHI EA '10, ACM, pp. 2813–2822.

[79] DREWES, H., AND SCHMIDT, A. The magic touch: Combining magic-pointing with a touch-sensitive mouse. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II* (Berlin, Heidelberg, 2009), INTERACT '09, Springer-Verlag, pp. 415–428.

[80] EDNEY, J. J. Human territories:" comment on functional properties". *Environment and Behavior 8*, 1 (1976), 31–47.

[81] EPPS, J., LICHMAN, S., AND WU, M. A study of hand shape use in tabletop gesture interaction. In *Ext. Abstracts CHI 2006* (2006), ACM Press, pp. 748–753.

[82] FALLMAN, D. Design-oriented human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2003), CHI '03, ACM, pp. 225–232.

[83] FAVALORA, G. E. Volumetric 3d displays and application infrastructure. *Computer 38* (August 2005), 37–44.

[84] FELFOLDY, G. L. Repetition effects in choice reaction time to multidimensional stimuli. *Perception & Psychophysics 15*, 3 (1974), 453–459.

[85] FISCHER, P. T., AND HORNECKER, E. Urban hci: Spatial aspects in the design of shared encounters for media facades. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, ACM, pp. 307–316.

[86] FISHER, E. R., BADAM, S. K., AND ELMQVIST, N. Designing peer-to-peer distributed user interfaces: Case studies on building distributed applications. *Int. J. Hum.-Comput. Stud. 72*, 1 (Jan. 2014), 100–110.

[87] FITTS, P. The information capacity of the human motor system in controlling the amplitude of movement. In *Journal of Experimental Psychology 47* (1954), pp. 381–391.

[88] FITTS, P. M., AND PETERSON, J. R. Information capacity of discrete motor responses. *J Exp Psychol 67* (Feb. 1964), 103–112.

[89] FLANDERS, M., AND SOECHTING, J. F. Kinematics of typing: parallel control of the two hands. *Journal of Neurophysiology 67*, 5 (1992), 1264–1274.

[90] FLEURY, C., DUVAL, T., GOURANTON, V., AND STEED, A. Evaluation of remote collaborative manipulation for scientific data analysis. In *Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2012), VRST '12, ACM, pp. 129–136.

[91] FOR STANDARDIZATION, I. O. ISO 9241-9: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9 - Requirements for non-keyboard input devices. Tech. rep., International Organization for Standardization, Geneva, 2002.

[92] FORLINES, C., AND BALAKRISHNAN, R. Evaluating tactile feedback and direct vs. indirect stylus input in pointing and crossing selection tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, ACM, pp. 1563–1572.

[93] FORLINES, C., VOGEL, D., AND BALAKRISHNAN, R. Hybridpointing: Fluid switching between absolute and relative pointing with a direct input device. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2006), UIST '06, ACM, pp. 211–220.

[94] FORLINES, C., WIGDOR, D., SHEN, C., AND BALAKRISHNAN, R. Direct-touch vs. mouse input for tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2007), CHI '07, ACM, pp. 647–656.

[95] FOUCAULT, C., MICAUX, M., BONNET, D., AND BEAUDOUIN-LAFON, M. Spad: A bimanual interaction technique for productivity applications on multi-touch tablets. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2014), CHI EA '14, ACM, pp. 1879–1884.

[96] FRANZ, E. A., ZELAZNIK, H. N., SWINNEN, S., AND WALTER, C. Spatial conceptual influences on the coordination of bimanual actions: when a dual task becomes a single task. *Journal of motor behavior 33*, 1 (2001), 103–112.

[97] FREES, S., KESSLER, G. D., AND KAY, E. Prism interaction for enhancing control in immersive virtual environments. *ACM Trans. Comput.-Hum. Interact. 14*, 1 (May 2007).

[98] FROEHLICH, B., HOCHSTRATE, J., SKUK, V., AND HUCKAUF, A. The globefish and the globemouse: two new six degree of freedom input devices for graphics applications. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems* (New York, NY, USA, 2006), ACM, pp. 191–199.

[99] FRÖHLICH, B., BLACH, R., STEFANI, O., HOCHSTRATE, J., HOFFMANN, J., KLÜGER, K., AND BUES, M. Implementing multi-viewer stereo displays. In *WSCG (Full Papers)* (2005), pp. 139–146.

[100] FRÖHLICH, B., TRAMBEREND, H., BEERS, A., AGRAWALA, M., AND BARAFF, D. Physically-based manipulation on the responsive workbench. In *Proceedings of the IEEE Virtual Reality 2000 Conference* (Washington, DC, USA, 2000), VR '00, IEEE Computer Society, pp. 5–.

[101] GARCÍA, A. S., MOLINA, J. P., GONZÁLEZ, P., MARTÍNEZ, D., AND MARTÍNEZ, J. An experimental study of collaborative interaction tasks supported by awareness and multimodal feedback. In *Proc. of VRCAI 2009* (2009), ACM Press, pp. 77–82.

[102] GARNER, W. R. *The processing of information and structure*. Lawrence Erlbaum, Oxford, 1974.

[103] GARNER, W. R., AND FELFOLDY, G. L. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology 1*, 3 (1970), 225–241.

[104] GAVER, W. What should we expect from research through design? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, ACM, pp. 937–946.

[105] GELFAND, I. M., AND LATASH, M. L. On the problem of adequate language in biology. *Progress in Motor Control: Structure-function relations in voluntary movements 2* (2002), 209.

[106] GENTNER, D. R., GRUDIN, J., AND CONWAY, E. Finger movements in transcription typing. Tech. rep., University of California, San Diego, Center for Human Information Processing, La Jolla, CA, USA, 1980.

[107] GENTRY, S., AND FERON, E. Human-human haptic collaboration in cyclical fitts' tasks. In *International Conference on Intelligent Robots and Systems* (2005).

[108] Goodale, M. A., Jakobson, L. S., and Servos, P. 2 - the visual pathways mediating perception and prehension. In *Hand and Brain*, A. M. W. H. R. Flanagan, Ed. Academic Press, San Diego, 1996, pp. 15 – 31.

[109] Graham, E. D., and MacKenzie, C. L. Physical versus virtual pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1996), CHI '96, ACM, pp. 292–299.

[110] Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., and Wang, M. Proxemic interactions: The new ubicomp? *interactions 18*, 1 (Jan. 2011), 42–50.

[111] Greenberg, S., and Marwood, D. Real time groupware as a distributed system: concurrency control and its effect on the interface. In *Proc. of CSCW 1994* (1994), CSCW '94, ACM Press, pp. 207–217.

[112] Gross, M., Würmlin, S., Naef, M., Lamboray, E., Spagno, C., Kunz, A., Koller-Meier, E., Svoboda, T., Van Gool, L., Lang, S., Strehlke, K., Moere, A. V., and Staadt, O. blue-c: a spatially immersive display and 3d video portal for telepresence. *ACM Trans. Graph. 22*, 3 (July 2003), 819–827.

[113] Grossman, T., and Balakrishnan, R. Pointing at trivariate targets in 3d environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2004), CHI '04, ACM, pp. 447–454.

[114] Grossman, T., and Balakrishnan, R. A probabilistic approach to modeling two-dimensional pointing. *ACM Trans. Comput.-Hum. Interact. 12*, 3 (Sept. 2005), 435–459.

[115] Guiard, Y. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. In *Journal of Motor Behavior* (1987), vol. 19, pp. 486–517.

[116] Guiard, Y. The kinematic chain as a model for human asymmetrical bimanual cooperation. *Advances in Psychology 55* (1988), 205–228.

[117] Guiard, Y. On fitts's and hooke's laws: Simple harmonic movement in upper-limb cyclical aiming. *Acta Psychologica 82*, 1-3 (1993), 139 – 159.

[118] Guiard, Y., and Beaudouin-Lafon, M. Target acquisition in multiscale electronic worlds. *International Journal of Human-Computer Studies 61*, 6 (2004), 875 – 905. Fitts' law 50 years later: applications and contributions from human-computer interaction.

[119] Guiard, Y., Beaudouin-Lafon, M., and Mottet, D. Navigation as multiscale pointing: Extending fitts' model to very high precision tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1999), CHI '99, ACM, pp. 450–457.

[120] Gutwin, C., and Greenberg, S. Design for individuals, design for groups: Tradeoffs between power and workspace awareness. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 1998), CSCW '98, ACM, pp. 207–216.

[121] Gutwin, C., and Greenberg, S. A descriptive framework of workspace awareness for real-time groupware. *Comput. Supported Coop. Work 11*, 3 (Nov. 2002), 411–446.

[122] Gutwin, C., and Greenberg, S. The importance of awareness for team cognition in distributed collaboration. In *Team Cognition: Understanding the Factors that Drive Process and Performance*, S. F. E. Salas, Ed. APA Press, 2004, pp. 177–201.

[123] Hachet, M., Guitton, P., and Reuter, P. The cat for efficient 2d and 3d interaction as an alternative to mouse adaptations. In *Proceedings of the ACM VRST'03* (2003), ACM, pp. 205–212.

[124] Haken, H., Kelso, J. S., and Bunz, H. A theoretical model of phase transitions in human hand movements. *Biological cybernetics 51*, 5 (1985), 347–356.

[125] Hall, E. T. *The Hidden Dimension*. Doubleday, 1966.

[126] Hancock, M., ten Cate, T., and Carpendale, S. Sticky tools: Full 6dof force-based interaction for multi-touch tables. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (New York, NY, USA, 2009), ITS '09, ACM, pp. 133–140.

[127] Hanson, F. Comparative studies of firefly pacemakers. *Federation proceedings 37*, 8 (June 1978), 2158–2164.

[128] Harrison, S., and Dourish, P. Re-place-ing space: The roles of place and space in collaborative systems. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 1996), CSCW '96, ACM, pp. 67–76.

[129] Hauber, J., Regenbrecht, H., Billinghurst, M., and Cockburn, A. Spatiality in videoconferencing: Trade-offs between efficiency and social presence. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (New York, NY, USA, 2006), CSCW '06, ACM, pp. 413–422.

[130] Hepp-Reymond, M., Huesler, E., and Maier, M. Precision grip in humans: temporal and spatial synergies. *Hand and brain* (1996), 33–68.

[131] Herrlich, M., Walther-Franks, B., and Malaka, R. Integrated rotation and translation for 3d manipulation on multi-touch interactive surfaces. In *Proceedings of the 11th International Conference on Smart Graphics* (Berlin, Heidelberg, 2011), SG'11, Springer-Verlag, pp. 146–154.

[132] HICK, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology 4*, 1 (1952), 11–26.

[133] HINCKLEY, K., BI, X., PAHUD, M., AND BUXTON, B. Informal information gathering techniques for active reading. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, ACM, pp. 1893–1896.

[134] HINCKLEY, K., CUTRELL, E., BATHICHE, S., AND MUSS, T. Quantitative analysis of scrolling techniques. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2002), ACM, pp. 65–72.

[135] HINCKLEY, K., CZERWINSKI, M., AND SINCLAIR, M. Interaction and modeling techniques for desktop two-handed input. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 1998), UIST '98, ACM, pp. 49–58.

[136] HINCKLEY, K., PAHUD, M., BENKO, H., IRANI, P., GUIMBRETIÈRE, F., GAVRILIU, M., CHEN, X. A., MATULIC, F., BUXTON, W., AND WILSON, A. Sensing techniques for tablet+stylus interaction. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2014), UIST '14, ACM, pp. 605–614.

[137] HINCKLEY, K., PAUSCH, R., GOBLE, J. C., AND KASSELL, N. F. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1994), CHI '94, ACM, pp. 452–458.

[138] HINCKLEY, K., PAUSCH, R., PROFFITT, D., PATTEN, J., AND KASSELL, N. Cooperative bimanual action. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1997), CHI '97, ACM, pp. 27–34.

[139] HINCKLEY, K., SINCLAIR, M., HANSON, E., SZELISKI, R., AND CONWAY, M. The videomouse: A camera-based multi-degree-of-freedom input device. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 1999), UIST '99, ACM, pp. 103–112.

[140] HINCKLEY, K., AND SONG, H. Sensor synaesthesia: Touch in motion, and motion in touch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2011), CHI '11, ACM, pp. 801–810.

[141] HINCKLEY, K., TULLIO, J., PAUSCH, R., PROFFITT, D., AND KASSELL, N. Usability analysis of 3d rotation techniques. In *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 1997), UIST '97, ACM, pp. 1–10.

[142] HINCKLEY, K., YATANI, K., PAHUD, M., CODDINGTON, N., RODENHOUSE, J., WILSON, A., BENKO, H., AND BUXTON, B. Manual deskterity: an exploration of simultaneous pen + touch direct input. In *Ext. Abstracts CHI 2010* (2010), ACM Press, pp. 2793–2802.

[143] HINCKLEY, K., YATANI, K., PAHUD, M., CODDINGTON, N., RODENHOUSE, J., WILSON, A., BENKO, H., AND BUXTON, B. Pen + touch = new tools. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2010), UIST '10, ACM, pp. 27–36.

[144] HINDMARSH, J., FRASER, M., HEATH, C., BENFORD, S., AND GREENHALGH, C. Object-focused interaction in collaborative virtual environments. *ACM Trans. Comput.-Hum. Interact. 7*, 4 (Dec. 2000), 477–509.

[145] HOFFMANN, E. R., DRURY, C. G., AND ROMANOWSKI, C. J. Performance in one-, two- and three-dimensional terminal aiming tasks. *Ergonomics 54*, 12 (2011), 1175–1185. PMID: 22103725.

[146] HOLZ, C., AND BAUDISCH, P. Fiberio: A touchscreen that senses fingerprints. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2013), UIST '13, ACM, pp. 41–50.

[147] HOPP, A. Using a single spatial light modulator for stereoscopic images of high color quality and resolution. In *IPT/EGVE* (2005), E. Kjems and R. Blach, Eds., Eurographics Association, pp. 15–22.

[148] HORNECKER, E., AND BUUR, J. Getting a grip on tangible interaction: A framework on physical space and social interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2006), CHI '06, ACM, pp. 437–446.

[149] HORNECKER, E., MARSHALL, P., DALTON, N. S., AND ROGERS, Y. Collaboration and interference: awareness with mice or touch input. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (New York, NY, USA, 2008), CSCW '08, ACM, pp. 167–176.

[150] HUA, H., BROWN, L., AND GAO, C. Scape: supporting stereoscopic collaboration in augmented and projective environments. *Computer Graphics and Applications, IEEE 24*, 1 (Jan-Feb 2004), 66–75.

[151] HUCKAUF, A., SPEED, A., KUNERT, A., HOCHSTRATE, J., AND FROEHLICH, B. Evaluation of 12-dof input devices for navigation and manipulation in virtual environments. In *INTERACT* (2005), pp. 601–614.

[152] HYMAN, R., AND WELL, A. Perceptual separability and spatial models. *Perception & Psychophysics 3*, 3 (1968), 161–165.

[153] IMAI, S., AND GARNER, W. Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology 69*, 6 (1965), 596.

[154] Isenberg, P., Tang, A., and Carpendale, S. An exploratory study of visual information analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, ACM, pp. 1217–1226.

[155] Ishii, H., and Kobayashi, M. Clearboard: a seamless medium for shared drawing and conversation with eye contact. In *Proc. of CHI 1992* (New York, NY, USA, 1992), ACM Press, pp. 525–532.

[156] Izadi, S., Brignull, H., Rodden, T., Rogers, Y., and Underwood, M. Dynamo: a public interactive surface supporting the cooperative sharing and exchange of media. In *Proceedings of the 16th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2003), UIST '03, ACM, pp. 159–168.

[157] Jacob, R., Silbert, R., Preston, M., and FARLANE, D. M. Integrality and Separability of Input Devices. In *Proc. of CHI '94* (1994), pp. 3–26.

[158] Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T., and Zigelbaum, J. Reality-based interaction: A framework for post-wimp interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, ACM, pp. 201–210.

[159] Jacobson, A., Panozzo, D., Glauser, O., Pradalier, C., Hilliges, O., and Sorkine-Hornung, O. Tangible and modular input device for character articulation. In *Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2014), UIST'14 Adjunct, ACM, pp. 45–46.

[160] Jakobsen, M. R., and Hornbæk, K. Up close and personal: Collaborative work on a high-resolution multitouch wall display. *ACM Trans. Comput.-Hum. Interact. 21*, 2 (Feb. 2014), 11:1–11:34.

[161] Jetter, H.-C., Gerken, J., Zöllner, M., Reiterer, H., and Milic-Frayling, N. Materializing the query with facet-streams: A hybrid surface for collaborative search on tabletops. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2011), CHI '11, ACM, pp. 3013–3022.

[162] Jones, A., Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M., and Debevec, P. Achieving eye contact in a one-to-many 3d video teleconferencing system. *ACM Trans. Graph. 28*, 3 (July 2009), 64:1–64:8.

[163] Jones, A., Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M., and Debevec, P. Achieving eye contact in a one-to-many 3d video teleconferencing system. *ACM Trans. Graph. 28* (July 2009), 64:1–64:8.

[164] JONES, A., LANG, M., FYFFE, G., YU, X., BUSCH, J., MCDOWALL, I., BOLAS, M., AND DEBEVEC, P. Headspin: a one-to-many 3d video teleconferencing system. In *ACM SIGGRAPH 2009 Emerging Technologies* (New York, NY, USA, 2009), SIGGRAPH '09, ACM, pp. 13:1–13:1.

[165] JONES, A., MCDOWALL, I., YAMADA, H., BOLAS, M., AND DEBEVEC, P. Rendering for an interactive 360 degree light field display. *ACM Trans. Graph. 26* (July 2007).

[166] JOTA, R., NG, A., DIETZ, P., AND WIGDOR, D. How fast is fast enough?: A study of the effects of latency in direct-touch pointing tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI '13, ACM, pp. 2291–2300.

[167] KABBASH, P., BUXTON, W., AND SELLEN, A. Two-handed input in a compound task. In *CHI '94: Conference companion on Human factors in computing systems* (New York, NY, USA, 1994), ACM, p. 230.

[168] KEELE, S. W. Movement control in skilled motor performance. *Psychological bulletin 70*, 6p1 (1968), 387.

[169] KELSO, J. Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 246*, 6 (1984), R1000–R1004.

[170] KELSO, J., DELCOLLE, J., AND SCHÖNER, G. Action-perception as a pattern formation process. *Attention and performance XIII 5* (1990), 139–169.

[171] KELSO, J., SOUTHARD, D., AND GOODMAN, D. On the nature of human interlimb coordination. *Science 203*, 4384 (1979), 1029–1031.

[172] KIM, M.-S., SEONG, J.-K., HYUN, D.-E., LEE, K.-H., AND CHOI, Y.-J. A physical 3d trackball. In *PG '01: Proceedings of the 9th Pacific Conference on Computer Graphics and Applications* (Washington, DC, USA, 2001), IEEE Computer Society, p. 134.

[173] KIN, K., MILLER, T., BOLLENSDORFF, B., DEROSE, T., HARTMANN, B., AND AGRAWALA, M. Eden: A professional multitouch tool for constructing virtual organic environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2011), CHI '11, ACM, pp. 1343–1352.

[174] KITAMURA, Y., KONISHI, T., YAMAMOTO, S., AND KISHINO, F. Interactive stereoscopic display for three or more users. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), SIGGRAPH '01, ACM, pp. 231–240.

[175] KITAMURA, Y., SAKURAI, S., YAMAGUCHI, T., FUKAZAWA, R., ITOH, Y., AND KISHINO, F. Multi-modal interface in multi-display environment for multi-users. In *Proceedings of the 13th International Conference on Human-Computer Interaction. Part II: Novel Interaction Methods and Techniques* (Berlin, Heidelberg, 2009), Springer-Verlag, pp. 66–74.

[176] KITAMURA, Y., YEE, A., KISHINO, F., AND SEIKA-CHO, H. A sophisticated manipulation aid in a virtual environment using dynamic constraints among object faces. *Presence 7* (1998), 460–477.

[177] KIYOKAWA, K., BILLINGHURST, M., HAYES, S. E., GUPTA, A., SANNOHE, Y., AND KATO, H. Communication behaviors of co-located users in collaborative ar interfaces. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality* (Washington, DC, USA, 2002), ISMAR '02, IEEE Computer Society, pp. 139–.

[178] KLAPP, S. Temporal compatibility in dual motor tasks ii: Simultaneous articulation and hand movements. *Memory & Cognition 9*, 4 (1981), 398–401.

[179] KLINKHAMMER, D., NITSCHE, M., SPECHT, M., AND REITERER, H. Adaptive personal territories for co-located tabletop interaction in a museum setting. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces* (New York, NY, USA, 2011), ITS '11, ACM, pp. 107–110.

[180] KNOBLICH, G., BUTTERFILL, S., AND SEBANZ, N. Chapter three - psychological research on joint action: Theory and data. In *Advances in Research and Theory*, B. H. Ross, Ed., vol. 54 of *Psychology of Learning and Motivation*. Academic Press, 2011, pp. 59 – 101.

[181] KNOEDEL, S., AND HACHET, M. Multi-touch rst in 2d and 3d spaces: Studying the impact of directness on user performance. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on* (March 2011), pp. 75–78.

[182] KORNBLUM, S., HASBROUCQ, T., AND OSMAN, A. Dimensional overlap: cognitive basis for stimulus-response compatibility–a model and taxonomy. *Psychological review 97*, 2 (1990), 253.

[183] KRÜGER, W., BOHN, C.-A., FRÖHLICH, B., SCHÜTH, H., STRAUSS, W., AND WESCHE, G. The responsive workbench: a virtual work environment. *Computer 28*, 7 (1995), 42–48.

[184] KUCK, R., WIND, J., RIEGE, K., AND BOGEN, M. Improving the avango vr/ar framework: Lessons learned. In *5th Workshop of the GI-VR/AR Group* (september 2008), Shaker, pp. 209–220.

[185] KULIK, A., DITTRICH, J., AND FROEHLICH, B. The hold-and-move gesture for multi-touch interfaces. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2012), MobileHCI '12, ACM, pp. 49–58.

[186] KULIK, A., HOCHSTRATE, J., KUNERT, A., AND FRÖHLICH, B. The influence of input device characteristics on spatial perception in desktop-based 3d applications. In *3D User Interfaces, 2009. 3DUI 2009. IEEE Symposium on* (March 2009), pp. 59–66.

[187] KULIK, A., KUNERT, A., BECK, S., REICHEL, R., BLACH, R., ZINK, A., AND FRÖHLICH, B. C1x6: a stereoscopic six-user display for co-located collaboration in shared virtual environments. *ACM Trans. Graph. 30*, 6 (Dec. 2011), 188:1–188:12.

[188] KULIK, A., KUNERT, A., LUX, C., AND FRÖHLICH, B. The pie slider: Combining advantages of the real and the virtual space. In *Smart Graphics* (2009), A. Butz, B. Fisher, M. Christie, A. Krüger, P. Olivier, and R. Therón, Eds., vol. 5531 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 93–104.

[189] KUNERT, A., KULIK, A., BECK, S., AND FRÖHLICH, B. Photoportals: Shared references in space and time. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (New York, NY, USA, 2014), CSCW '14, ACM, pp. 1388–1399.

[190] KUNERT, A., KULIK, A., HUCKAUF, A., AND FRÖHLICH, B. A comparison of tracking- and controller-based input for complex bimanual interaction in virtual environments. In *Proceedings of the 13th Eurographics Conference on Virtual Environments* (Aire-la-Ville, Switzerland, Switzerland, 2007), EGVE'07, Eurographics Association, pp. 43–52.

[191] KUNZ, A. M., AND SPAGNO, C. P. Technical system for collaborative work. In *Proceedings of the Eurographics Workshop on Virtual Environments 2002* (2002), EGVE '02, pp. 73–80.

[192] LAKATOS, I. Falsification and the methodology of scientific research programmes. In *Can Theories be Refuted?* Springer, 1976, pp. 205–259.

[193] LANK, E., RUIZ, J., AND COWAN, W. Concurrent bimanual stylus interaction: a study of non-preferred hand mode manipulation. In *Proc. Graphics Interface 2006* (2006), CIPS, pp. 17–24.

[194] LARGE, E. W., AND JONES, M. R. The dynamics of attending: how people track time-varying events. *Psychological review 106*, 1 (1999), 119.

[195] LASHLEY, K. S. Integrative functions of the cerebral cortex. *Physiological Reviews 13*, 1 (1933), 1–42.

[196] LATASH, M. *Neurophysiological basis of movement*. Human Kinetics, Champaign, IL, 2008.

[197] LATASH, M. L. 2 - elements of history. In *Fundamentals of Motor Control*, M. L. Latash, Ed. Academic Press, San Diego, 2012, pp. 5 – 24.

[198] LATASH, M. L. 9 - coordination. In *Fundamentals of Motor Control*, M. L. Latash, Ed. Academic Press, San Diego, 2012, pp. 149 – 170.

[199] LATASH, M. L., SCHOLZ, J. P., AND SCHONER, G. Toward a new theory of motor synergies. *MOTOR CONTROL-CHAMPAIGN- 11*, 3 (2007), 276.

[200] LATULIPE, C., KAPLAN, C. S., AND CLARKE, C. L. A. Bimanual and unimanual image alignment: An evaluation of mouse-based techniques. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2005), UIST '05, ACM, pp. 123–131.

[201] LATULIPE, C., MANN, S., KAPLAN, C. S., AND CLARKE, C. L. A. symspline: Symmetric two-handed spline manipulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2006), CHI '06, ACM, pp. 349–358.

[202] LAVIOLA JR, J. J., FORSBERG, A. S., HUFFMAN, J., AND BRAGDON, A. The influence of head tracking and stereo on user performance with non-isomorphic 3d rotation. In *IPT/EGVE* (2008), pp. 111–118.

[203] LEGANCHUCK, A., ZHAI, S., AND BUXTON, W. Manual and Cognitive Benefits of two handed Input: an Experimental Study. In *Proc. of CHI '99* (1999), pp. 326–359.

[204] LEGANCHUK, A., ZHAI, S., AND BUXTON, W. Manual and cognitive benefits of two-handed input: an experimental study. *ACM Trans. Comput.-Hum. Interact. 5*, 4 (1998), 326–359.

[205] LI, Y., HINCKLEY, K., GUAN, Z., AND LANDAY, J. A. Experimental analysis of mode switching techniques in pen-based user interfaces. In *Proc. CHI 2005* (2005), ACM Press, pp. 461–470.

[206] LIU, J., AU, O. K.-C., FU, H., AND TAI, C.-L. Two-Finger Gestures for 6DOF Manipulation of 3D Objects. *Computer Graphics Forum 31*, 7pt1 (Sept. 2012), 2047–2055.

[207] LIU, L., MARTENS, J.-B., AND VAN LIERE, R. Revisiting path steering for 3d manipulation tasks. *Int. J. Hum.-Comput. Stud. 69*, 3 (Mar. 2011), 170–181.

[208] LLAMAS, I., POWELL, A., ROSSIGNAC, J., AND SHAW, C. D. Bender: A virtual ribbon for deforming 3d shapes in biomedical and styling applications. In *Proceedings of the 2005 ACM Symposium on Solid and Physical Modeling* (New York, NY, USA, 2005), SPM '05, ACM, pp. 89–99.

[209] LOGAN, G. D., AND CRUMP, M. J. The left hand doesn't know what the right hand is doing the disruptive effects of attention to the hands in skilled typewriting. *Psychological Science 20*, 10 (2009), 1296–1300.

[210] LOGAN, G. D., AND CRUMP, M. J. Hierarchical control of cognitive processes: The case for skilled typewriting. *Psychology of Learning and Motivation-Advances in Research and Theory 54* (2011), 1.

[211] LU, H., AND LI, Y. Gesture avatar: a technique for operating mobile user interfaces using gestures. In *Proc. CHI 2011* (2011), ACM Press, pp. 207–216.

[212] MACKENZIE, I. S. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact. 7*, 1 (Mar. 1992), 91–139.

[213] MACKENZIE, I. S., AND BUXTON, W. Extending fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1992), CHI '92, ACM, pp. 219–226.

[214] MACKENZIE, I. S., AND ISOKOSKI, P. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, ACM, pp. 1633–1636.

[215] MACKENZIE, I. S., SELLEN, A., AND BUXTON, W. A. S. A comparison of input devices in element pointing and dragging tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1991), CHI '91, ACM, pp. 161–166.

[216] MACKENZIE, I. S., SOUKOREFF, R. W., AND PAL, C. A two-ball mouse affords three degrees of freedom. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 1997), CHI EA '97, ACM, pp. 303–304.

[217] MACKENZIE, I. S., AND WARE, C. Lag as a determinant of human performance in interactive systems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (New York, NY, USA, 1993), CHI '93, ACM, pp. 488–493.

[218] MAIMONE, A., AND FUCHS, H. Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras. In *Proc. of ISMAR 2011* (Washington, DC, USA, 2011), IEEE Computer Society, pp. 137–146.

[219] MAKSAKOV, E., BOOTH, K. S., AND HAWKEY, K. Whale tank virtual reality. In *Proceedings of Graphics Interface 2010* (2010), GI '10, Canadian Information Processing Society, pp. 185–192.

[220] MARQUARDT, N., KIEMER, J., AND GREENBERG, S. What caused that touch?: Expressive interaction with a surface through fiduciary-tagged gloves. In *ACM International Conference on Interactive Tabletops and Surfaces* (New York, NY, USA, 2010), ITS '10, ACM, pp. 139–142.

[221] MARTINET, A., CASIEZ, G., AND GRISONI, L. Integrality and separability of multitouch interaction techniques in 3d manipulation tasks. *Visualization and Computer Graphics, IEEE Transactions on 18*, 3 (March 2012), 369–380.

[222] MASLIAH, M. R., AND MILGRAM, P. Measuring the allocation of control in a 6 degree-of-freedom docking experiment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2000), CHI '00, ACM, pp. 25–32.

[223] MATEJKA, J., GROSSMAN, T., LO, J., AND FITZMAURICE, G. The design and evaluation of multi-finger mouse emulation techniques. In *Proc. CHI 2009* (2009), ACM Press, pp. 1073–1082.

[224] MATSUSHITA, N., AYATSUKA, Y., AND REKIMOTO, J. Dual touch: a two-handed interface for pen-based pdas. In *Proc. UIST 2000* (2000), ACM Press, pp. 211–212.

[225] MATTESSICH, P. W., AND MONSEY, B. R. *Collaboration: what makes it work. A review of research literature on factors influencing successful collaboration.* ERIC, 1992.

[226] MCCALLUM, D. C., AND IRANI, P. Arc-pad: Absolute+relative cursor positioning for large displays with a mobile touchscreen. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2009), UIST '09, ACM, pp. 153–156.

[227] MCGRATH, W., BOWMAN, B., MCCALLUM, D., HINCAPIÉ-RAMOS, J. D., ELMQVIST, N., AND IRANI, P. Branch-explore-merge: facilitating real-time revision control in collaborative visual exploration. In *Proc. of ITS 2012* (2012), ITS '12, ACM, pp. 235–244.

[228] MECHSNER, F., KERZEL, D., KNOBLICH, G., AND PRINZ, W. Perceptual basis of bimanual coordination. *Nature 414*, 6859 (2001), 69–73.

[229] MEEHAN, M., RAZZAQUE, S., WHITTON, M. C., AND BROOKS, F. P. Effect of latency on presence in stressful virtual environments. In *virtual reality, 2003. Proceedings. IEEE* (2003), IEEE, pp. 141–148.

[230] MEYER, D. E., ABRAMS, R. A., KORNBLUM, S., WRIGHT, C. E., AND SMITH, J. E. K. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological Review 95* (1988), 340–370.

[231] MINE, M. R., BROOKS, JR., F. P., AND SEQUIN, C. H. Moving objects in space: Exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1997), SIGGRAPH '97, ACM Press/Addison-Wesley Publishing Co., pp. 19–26.

[232] MORIKAWA, O., AND MAESAKO, T. Hypermirror: toward pleasant-to-use video mediated communication system. In *Proc of CSCW 1998* (New York, NY, USA, 1998), ACM Press, pp. 149–158.

[233] MORRIS, M. R., CASSANEGO, A., PAEPCKE, A., WINOGRAD, T., PIPER, A. M., AND HUANG, A. Mediating group dynamics through tabletop interface design. *IEEE Comput. Graph. Appl. 26*, 5 (Sept. 2006), 65–73.

[234] MORRIS, M. R., HUANG, A., PAEPCKE, A., AND WINOGRAD, T. Cooperative gestures: Multi-user gestural interactions for co-located groupware. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2006), CHI '06, ACM, pp. 1201–1210.

[235] MORRIS, M. R., RYALL, K., SHEN, C., FORLINES, C., AND VERNIER, F. Beyond "social protocols": multi-user coordination policies for co-located groupware. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (New York, NY, USA, 2004), CSCW '04, ACM, pp. 262–265.

[236] MOSCOVICH, T. Multi-touch interaction. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2006), CHI EA '06, ACM, pp. 1775–1778.

[237] MOTTET, D., AND BOOTSMA, R. J. The dynamics of goal-directed rhythmical aiming. *Biological cybernetics 80*, 4 (1999), 235–245.

[238] MOTTET, D., GUIARD, Y., FERRAND, T., AND BOOTSMA, R. J. Two-handed performance of a rhythmical fitts task by individuals and dyads. *Journal of Experimental Psychology: Human Perception and Performance 27*, 6 (2001), 1275.

[239] MUHLBACH, L., BOCKER, M., AND PRUSSOG, A. Telepresence in videocommunications: a study on stereoscopy and individual eye contact. *Human Factors 37*, 2 (1995), 290–305.

[240] MULDER, J. D., AND BOSCHKER, B. R. A modular system for collaborative desktop vr/ar with a shared workspace. *Virtual Reality Conference, IEEE* (2004), 75–83.

[241] MURATA, A., AND IWASE, H. Extending fitts' law to a three-dimensional pointing task. *Human Movement Science 20* (2001), 791–805.

[242] NACENTA, M. A., BAUDISCH, P., BENKO, H., AND WILSON, A. Separability of spatial manipulations in multi-touch interfaces. In *Proceedings of Graphics Interface 2009* (Toronto, Ont., Canada, Canada, 2009), GI '09, Canadian Information Processing Society, pp. 175–182.

[243] NACENTA, M. A., JAKOBSEN, M. R., DAUTRICHE, R., HINRICHS, U., DÖRK, M., HABER, J., AND CARPENDALE, S. The lunchtable: A multi-user, multi-display system for information sharing in casual group interactions. In *Proceedings of the 2012 International Symposium on Pervasive Displays* (New York, NY, USA, 2012), PerDis '12, ACM, pp. 18:1–18:6.

[244] NANCEL, M., CHAPUIS, O., PIETRIGA, E., YANG, X.-D., IRANI, P. P., AND BEAUDOUIN-LAFON, M. High-precision pointing on large wall displays using small handheld devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI '13, ACM, pp. 831–840.

[245] NGUYEN, D., AND CANNY, J. Multiview: spatially faithful group video conferencing. In *Proc. of CHI 2005* (New York, NY, USA, 2005), ACM Press, pp. 799–808.

[246] NIEUWENHUIZEN, K., MARTENS, J.-B., LIU, L., AND VAN LIERE, R. Insights from dividing 3d goal-directed movements into meaningful phases. *Computer Graphics and Applications, IEEE 29*, 6 (Nov 2009), 44–53.

[247] NITZSCHE, N., HANEBECK, U. D., AND SCHMIDT, G. Motion compression for telepresent walking in large target environments. *Presence: Teleoper. Virtual Environ. 13* (February 2004), 44–60.

[248] NORMAN, D. The next ui breakthrough: Command lines. *interactions 14*, 3 (May 2007), 44–45.

[249] OLWAL, A., FEINER, S., AND HEYMAN, S. Rubbing and tapping for precise and rapid selection on touch-screen displays. In *Proc. CHI 2008* (2008), ACM Press, pp. 295–304.

[250] OSAWA, N., AND REN, X. Gearbox widget for fine adjustments by hand motion. In *Proceedings of the Workshop on Virtual Environments 2003* (New York, NY, USA, 2003), EGVE '03, ACM, pp. 313–314.

[251] OSAWA, N., AND REN, X. A study on approximate and fine adjustments by hand motion in an immersive environment(human-interface basics,<special issue> game programming). *IPSJ Journal 48*, 11 (nov 2007), 3568–3576.

[252] OTTO, O., ROBERTS, D., AND WOLFF, R. A review on effective closely-coupled collaboration using immersive cve's. In *Proceedings of the ACM VRCIA '06* (2006), ACM, pp. 145–154.

[253] OULASVIRTA, A., ROOS, T., MODIG, A., AND LEPPÄNEN, L. Information capacity of full-body movements. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI '13, ACM, pp. 1289–1298.

[254] OVIATT, S. Multimodal interactive maps: Designing for human performance. *Hum.-Comput. Interact. 12*, 1 (Mar. 1997), 93–129.

[255] OVIATT, S. Ten myths of multimodal interaction. *Commun. ACM 42*, 11 (Nov. 1999), 74–81.

[256] OVIATT, S., COHEN, P., WU, L., VERGO, J., DUNCAN, L., SUHM, B., BERS, J., HOLZMAN, T., WINOGRAD, T., LANDAY, J., LARSON, J., AND FERRO, D. Designing the user interface for multimodal speech and pen-based gesture applications: State-of-the-art systems and future research directions. *Hum.-Comput. Interact. 15*, 4 (Dec. 2000), 263–322.

[257] PARSONS, L. M. Inability to reason about an object's orientation using an axis and angle of rotation, 1995.

[258] PATLA, E. A., PRENTICE, D. S., RIETDYK, S., ALLARD, F., AND MARTIN, C. What guides the selection of alternate foot placement during locomotion in humans. *Experimental Brain Research 128*, 4 (1999), 441–450.

[259] PAULIGNAN, Y., AND JEANNEROD, M. 13 - prehension movements: The visuomotor channels hypothesis revisited. In *Hand and Brain*, A. M. W. H. R. Flanagan, Ed. Academic Press, San Diego, 1996, pp. 265 – 282.

[260] PAVLOVYCH, A., AND STUERZLINGER, W. The tradeoff between spatial jitter and latency in pointing tasks. In *Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (New York, NY, USA, 2009), EICS '09, ACM, pp. 187–196.

[261] PEDERSEN, E. W., AND HORNBÆK, K. An experimental comparison of touch interaction on vertical and horizontal surfaces. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (New York, NY, USA, 2012), NordiCHI '12, ACM, pp. 370–379.

[262] PELTONEN, P., KURVINEN, E., SALOVAARA, A., JACUCCI, G., ILMONEN, T., EVANS, J., OULASVIRTA, A., AND SAARIKKO, P. It's mine, don't touch!: interactions at a large multi-touch display in a city centre. In *Proceeding of CHI08* (New York, NY, USA, 2008), ACM, pp. 1285–1294.

[263] PETERS, M. Simultaneous performance of two motor activities: The factor of timing. *Neuropsychologia 15*, 3 (1977), 461–465.

[264] PETERS, M. Footedness: asymmetries in foot preference and skill and neuropsychological assessment of foot movement. *Psychological Bulletin 103*, 2 (1988), 179.

[265] PETIT, B., LESAGE, J.-D., MENIER, C., ALLARD, J., FRANCO, J.-S., RAFFIN, B., BOYER, E., AND FAURE, F. Multicamera real-time 3d modeling for telepresence and remote collaboration. *International Journal of Digital Multimedia Broadcasting 2010* (2009), 247108–12.

[266] PFEUFFER, K., ALEXANDER, J., AND GELLERSEN, H. Partially-indirect bimanual input with gaze, pen, and touch for pan, zoom, and ink interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2016), CHI '16, ACM, pp. 2845–2856.

[267] PIERCE, J. S., STEARNS, B. C., AND PAUSCH, R. Voodoo dolls: seamless interaction at multiple scales in virtual environments. In *Proc of I3D 1999* (1999), ACM Press, pp. 141–145.

[268] PINELLE, D., BARJAWI, M., NACENTA, M., AND MANDRYK, R. An evaluation of coordination techniques for protecting objects and territories in tabletop groupware. In *Proc. of CHI 2009* (2009), CHI '09, ACM, pp. 2129–2138.

[269] PINHO, M. S., BOWMAN, D. A., AND FREITAS, C. M. Cooperative object manipulation in immersive virtual environments: framework and techniques. In *Proc. of VRST 2002* (2002), VRST '02, ACM Press, pp. 171–178.

[270] PIRCHHEIM, C., WALDNER, M., AND SCHMALSTIEG, D. Deskotheque: Improved spatial awareness in multi-display environments. In *Virtual Reality Conference, 2009. VR 2009. IEEE* (march 2009), pp. 123 –126.

[271] POPPER, K. R. The logic of scientific discovery. *London: Hutchinson* (1959).

[272] POSTON, T., AND SERRA, L. The virtual workbench: dextrous vr. In *Proceedings of the conference on Virtual reality software and technology* (River Edge, NJ, USA, 1994), World Scientific Publishing Co., Inc., pp. 111–121.

[273] POTTS, B. C., MELARA, R. D., AND MARKS, L. E. Circle size and diameter tilt: A new look at integrality and separability. *Perception & psychophysics 60*, 1 (1998), 101–112.

[274] POUPYREV, I., WEGHORST, S., AND FELS, S. Non-isomorphic 3d rotational techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2000), CHI '00, ACM, pp. 540–547.

[275] PROCTOR, R. W., AND VU, K.-P. L. *Stimulus-Response Compatibility Principles: Data, Theory, and Application*. CRC Press, Boca Raton, Florida, USA, 2006.

[276] RÄDLE, R., JETTER, H.-C., MARQUARDT, N., REITERER, H., AND ROGERS, Y. Huddlelamp: Spatially-aware mobile displays for ad-hoc around-the-table collaboration. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (New York, NY, USA, 2014), ITS '14, ACM, pp. 45–54.

[277] RAGHUPATHY, P. B., AND BORST, C. W. Investigation of secondary views in a multimodal vr environment: 3d lenses, windows, and mirrors. In *Advances in Visual Computing*. Springer, 2011, pp. 180–189.

[278] RAMOS, G., AND BALAKRISHNAN, R. Zliding: Fluid zooming and sliding for high precision parameter manipulation. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2005), UIST '05, ACM, pp. 143–152.

[279] RAZZAQUE, S., SWAPP, D., SLATER, M., WHITTON, M. C., AND STEED, A. Redirected walking in place. In *EGVE '02: Proceedings of the Eurographics Workshop on Virtual Environments 2002* (2002), Eurographics Association, pp. 123–130.

[280] REISMAN, J. L., DAVIDSON, P. L., AND HAN, J. Y. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2009), UIST '09, ACM, pp. 69–78.

[281] REUTER, P., RIVIERE, G., COUTURE, N., SORRAING, N., ESPINASSE, L., AND VERGNIEUX, R. ArcheoTUI - A Tangible User Interface for the Virtual Reassembly of Fractured Archeological Objects. In *The 8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST'07)* (Brighton, United Kingdom, Nov. 2007), Eurographics Association, pp. 15–22. 8 pages.

[282] RIEGE, K., HOLTKAMPER, T., WESCHE, G., AND FROHLICH, B. The bent pick ray: An extended pointing technique for multi-user interaction. In *Proc. of 3D UI 2006* (2006), 3DUI '06, IEEE Computer Society, pp. 62–65.

[283] RILEY, M. A., RICHARDSON, M. J., SHOCKLEY, K., AND RAMENZONI, V. C. Interpersonal synergies. *Frontiers in psychology 2* (2011).

[284] RITTEL, H. W. J., AND WEBBER, M. Dilemmas in general theory of planning. *Policy Sciences 4* (1973), 155–169.

[285] ROBERTS, D., WOLFF, R., OTTO, O., AND STEED, A. Constructing a gazebo: Supporting teamwork in a tightly coupled, distributed task in virtual reality. *Presence: Teleoper. Virtual Environ. 12*, 6 (Dec. 2003), 644–657.

[286] ROSENBAUM, D. A. Chapter 2 - core problems. In *Human Motor Control (Second Edition)*, D. A. Rosenbaum, Ed., second edition ed. Academic Press, San Diego, 2010, pp. 11 – 41.

[287] ROSENBAUM, D. A. Chapter 5 - walking. In *Human Motor Control (Second Edition)*, D. A. Rosenbaum, Ed., second edition ed. Academic Press, San Diego, 2010, pp. 134A – 171.

[288] ROSENBAUM, D. A. Chapter 7 - reaching and grasping. In *Human Motor Control (Second Edition)*, D. A. Rosenbaum, Ed., second edition ed. Academic Press, San Diego, 2010, pp. 211 – 250.

[289] ROSENBAUM, D. A. Chapter 9 - keyboarding. In *Human Motor Control (Second Edition)*, D. A. Rosenbaum, Ed., second edition ed. Academic Press, San Diego, 2010, pp. 277 – 321.

[290] ROSENBAUM, D. A., MEULENBROEK, R. G., AND VAUGHAN, J. 9 - three approaches to the degrees of freedom problem in reaching. In *Hand and Brain*, A. M. W. H. R. Flanagan, Ed. Academic Press, San Diego, 1996, pp. 169 – 185.

[291] ROTH, V., AND TURNER, T. Bezel swipe: conflict-free scrolling and multiple selection on mobile touch screen devices. In *Proc. CHI 2009* (2009), ACM Press, pp. 1523–1526.

[292] ROUDAUT, A., LECOLINET, E., AND GUIARD, Y. Microrolls: expanding touch-screen input vocabulary by distinguishing rolls vs. slides of the thumb. In *Proc. CHI 2009* (2009), ACM Press, pp. 927–936.

[293] RUDDLE, R. A., SAVAGE, J. C. D., AND JONES, D. M. Symmetric and asymmetric action integration during cooperative object manipulation in virtual environments. *ACM Trans. Comput.-Hum. Interact. 9*, 4 (Dec. 2002), 285–308.

[294] RUSSELL, D. M., DREWS, C., AND SUE, A. Social aspects of using large public interactive displays for collaboration. In *Proceedings of the 4th International Conference on Ubiquitous Computing* (London, UK, UK, 2002), UbiComp '02, Springer-Verlag, pp. 229–236.

[295] RYALL, K., FORLINES, C., SHEN, C., AND MORRIS, M. R. Exploring the effects of group size and table size on interactions with tabletop shared-display groupware. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (New York, NY, USA, 2004), CSCW '04, ACM, pp. 284–293.

[296] SALZMAN, ., STACHNIAK, S., AND STUERZLINGER, W. Unconstrained vs. constrained 3d scene manipulation. In *In 8th IFIP International Conference on Engineering for Human-Computer Interaction* (2001), p. 2001.

[297] SALZMANN, H., JACOBS, J., AND FROEHLICH, B. Collaborative interaction in co-located two-user scenarios. In *Proceedings of the 15th Joint Virtual Reality Eurographics Conference on Virtual Environments* (Aire-la-Ville, Switzerland, Switzerland, 2009), JVRC'09, Eurographics Association, pp. 85–92.

[298] SASANGOHAR, F., MACKENZIE, I., S., AND SCOTT, STACEY, D. Evaluation of mouse and touch input for a tabletop display using fitts' reciprocal tapping task. In *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society, HFES 2009* (Santa Monica, CA, USA, 2009), HFES 2009, Human Factors and Ergonomics Society, pp. 839–843.

[299] SAWYER, K. *Group genius: The creative power of collaboration*. Basic Books, 2008.

[300] SCHMALSTIEG, D., FUHRMANN, A., HESINA, G., SZALAVÁRI, Z., ENCARNAÇÃO, L. M., GERVAUTZ, M., AND PURGATHOFER, W. The studierstube augmented reality project. *Presence: Teleoper. Virtual Environ. 11* (February 2002), 33–54.

[301] SCHMIDT, R., BIENVENU, M., FITZPATRICK, P., AND AMAZEEN, P. A comparison of intra-and interpersonal interlimb coordination: coordination breakdowns and coupling strength. *Journal of Experimental Psychology: Human Perception and Performance 24*, 3 (1998), 884.

[302] SCHMIDT, R., SINGH, K., AND BALAKRISHNAN, R. Sketching and composing widgets for 3d manipulation. *Computer Graphics Forum* (2008), 301–310.

[303] SCHMIDT, R. C., AND RICHARDSON, M. J. Dynamics of interpersonal coordination. In *Coordination: Neural, behavioral and social dynamics*. Springer, 2008, pp. 281–308.

[304] SCHNEIDER, D. W., AND LOGAN, G. D. Hierarchical control of cognitive processes: Switching tasks in sequences. *Journal of Experimental Psychology: General 135*, 4 (2006), 623.

[305] SCHNELLE-WALKA, D., AND DÖWELING, S. Speech augmented multitouch interaction patterns. In *Proceedings of the 16th European Conference on Pattern Languages of Programs* (New York, NY, USA, 2012), EuroPLoP '11, ACM, pp. 8:1–8:23.

[306] SCHOLZ, J. P., AND SCHÖNER, G. The uncontrolled manifold concept: identifying control variables for a functional task. *Experimental brain research 126*, 3 (1999), 289–306.

[307] SCHÖNER, G., HAKEN, H., AND KELSO, J. A stochastic theory of phase transitions in human hand movement. *Biological cybernetics 53*, 4 (1986), 247–257.

[308] SCHREER, O., FELDMANN, I., ATZPADIN, N., EISERT, P., KAUFF, P., AND BELT, H. 3dpresence -a system concept for multi-user and multi-party immersive 3d videoconferencing. In *Proc. of CVMP 2008* (nov. 2008), pp. 1–8.

[309] SCHROEDER, R., STEED, A., AXELSSON, A.-S., HELDAL, I., ABELIN, Å., WIDESTRÖM, J., NILSSON, A., AND SLATER, M. Collaborating in networked immersive spaces: as good as being there together? *Computers & Graphics 25*, 5 (2001), 781–788.

[310] SCOTT, S. D., GRANT, K. D., AND MANDRYK, R. L. System guidelines for co-located, collaborative work on a tabletop display. In *Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work* (Norwell, MA, USA, 2003), Kluwer Academic Publishers, pp. 159–178.

[311] SCOTT, S. D., SHEELAGH, M., CARPENDALE, T., AND INKPEN, K. M. Territoriality in collaborative tabletop workspaces. In *Proc. of CSCW 2004* (2004), CSCW '04, ACM Press, pp. 294–303.

[312] SHAFFER, L. Latency mechanisms in transcription. *Attention and performance 4* (1973), 435–446.

[313] SHARMA, A., MADHVANATH, S., SHEKHAWAT, A., AND BILLINGHURST, M. Mozart: A multimodal interface for conceptual 3d modeling. In *Proceedings of the 13th International Conference on Multimodal Interfaces* (New York, NY, USA, 2011), ICMI '11, ACM, pp. 307–310.

[314] SHAW, C. D., AND GREEN, M. Thred: a two-handed design system. *Multimedia Systems 5*, 2 (1997), 126–139.

[315] SHEPP, B. E. On perceiving objects: Holistic versus featural properties. In *Object Perception: Structure and Process*, B. Shepp and S. Ballesteros, Eds. Lawrence Erlbaum Associates, Inc, 1989.

[316] SHNEIDERMAN, B. Direct manipulation: A step beyond programming languages. *IEEE Computer 16* (1983), 57–69.

[317] SHOEMAKE, K. Arcball: a user interface for specifying three-dimensional orientation using a mouse. In *Proceedings of the conference on Graphics interface '92* (San Francisco, CA, USA, 1992), Morgan Kaufmann Publishers Inc., pp. 151–156.

[318] SIIO, I., AND TSUJITA, H. Mobile interaction using paperweight metaphor. In *Proc. UIST 2006* (2006), ACM Press, pp. 111–114.

[319] SMITH, G., SALZMAN, T., AND STUERZLINGER, W. 3d scene manipulation with 2d devices and constraints. In *GRIN'01: No description on Graphics interface 2001* (Toronto, Ont., Canada, Canada, 2001), Canadian Information Processing Society, pp. 135–142.

[320] SOLOVEY, E., SCHERMERHORN, P., SCHEUTZ, M., SASSAROLI, A., FANTINI, S., AND JACOB, R. Brainput: Enhancing interactive systems with streaming fnirs brain input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, ACM, pp. 2193–2202.

[321] SOUKOREFF, R. W., AND MACKENZIE, I. S. Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *Int. J. Hum.-Comput. Stud. 61*, 6 (Dec. 2004), 751–789.

[322] STEINICKE, F., BRUDER, G., JERALD, J., FRENZ, H., AND LAPPE, M. Analyses of human sensitivity to redirected walking. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology* (2008), VRST '08, pp. 149–156.

[323] STELLMACH, S., AND DACHSELT, R. Look & touch: Gaze-supported target acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, ACM, pp. 2981–2990.

[324] STELLMACH, S., AND DACHSELT, R. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI '13, ACM, pp. 285–294.

[325] STOAKLEY, R., CONWAY, M. J., AND PAUSCH, R. Virtual reality on a wim: interactive worlds in miniature. In *Proc. of CHI 1995* (New York, NY, USA, 1995), ACM Press/Addison-Wesley Publishing Co., pp. 265–272.

[326] STOELEN, M. F., AND AKIN, D. L. Assessment of fitts' law for quantifying combined rotational and translational movements. *Human Factors: The Journal of the Human Factors and Ergonomics Society 52*, 1 (2010), 63–77.

[327] STOEV, S. L., AND SCHMALSTIEG, D. Application and taxonomy of through-the-lens techniques. In *Proc. of VRST 2002* (2002), ACM Press, pp. 57–64.

[328] STREITZ, N. A., GEISSLER, J., HOLMER, T., KONOMI, S., MÜLLER-TOMFELDE, C., REISCHL, W., REXROTH, P., SEITZ, P., AND STEINMETZ, R. i-land: An interactive landscape for creativity and innovation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1999), CHI '99, ACM, pp. 120–127.

[329] STURMAN, D. J., ZELTZER, D., AND PIEPER, S. Hands-on interaction with virtual environments. In *Proceedings of the 2Nd Annual ACM SIGGRAPH Symposium on User Interface Software and Technology* (New York, NY, USA, 1989), UIST '89, ACM, pp. 19–24.

[330] STÜRMER, B., ASCHERSLEBEN, G., AND PRINZ, W. Correspondence effects with manual gestures and postures: a study of imitation. *Journal of Experimental Psychology: Human Perception and Performance 26*, 6 (2000), 1746.

[331] SUTHERLAND, I. E. Sketchpad: A man-machine graphical communication system. In *Proceedings of the May 21-23, 1963, Spring Joint Computer Conference* (New York, NY, USA, 1963), AFIPS '63 (Spring), ACM, pp. 329–346.

[332] TANDLER, P., PRANTE, T., MÜLLER-TOMFELDE, C., STREITZ, N., AND STEINMETZ, R. Connectables: dynamic coupling of displays for the flexible creation of shared workspaces. In *Proceedings of the 14th annual ACM symposium on User interface software and technology* (New York, NY, USA, 2001), UIST '01, ACM, pp. 11–20.

[333] TANG, A., TORY, M., PO, B., NEUMANN, P., AND CARPENDALE, S. Collaborative coupling over tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2006), CHI '06, ACM, pp. 1181–1190.

[334] TEATHER, R., AND STUERZLINGER, W. Pointing at 3d targets in a stereo head-tracked virtual environment. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on* (March 2011), pp. 87–94.

[335] TEATHER, R. J., AND STUERZLINGER, W. Guidelines for 3d positioning techniques. In *Proceedings of the 2007 Conference on Future Play* (New York, NY, USA, 2007), Future Play '07, ACM, pp. 61–68.

[336] TOBIASZ, M., ISENBERG, P., AND CARPENDALE, S. Lark: Coordinating co-located collaboration with information visualization. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (Nov. 2009), 1065–1072.

[337] TODOROV, E. Optimality principles in sensorimotor control. *Nature neuroscience 7*, 9 (2004), 907–915.

[338] TRAN, J. J., TREWIN, S., SWART, C., JOHN, B. E., AND THOMAS, J. C. Exploring pinch and spread gestures on mobile devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2013), MobileHCI '13, ACM, pp. 151–160.

[339] TSE, E., HANCOCK, M., AND GREENBERG, S. Speech-filtered bubble ray: Improving target acquisition on display walls. In *Proceedings of the 9th International Conference on Multimodal Interfaces* (New York, NY, USA, 2007), ICMI '07, ACM, pp. 307–314.

[340] TSE, E., HISTON, J., SCOTT, S. D., AND GREENBERG, S. Avoiding interference: how people use spatial separation and partitioning in sdg workspaces. In *Proc. of CSCW 2004* (2004), ACM Press, pp. 252–261.

[341] TSE, E., SHEN, C., GREENBERG, S., AND FORLINES, C. Enabling interaction with single user applications through speech and gestures on a multi-user tabletop. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (New York, NY, USA, 2006), AVI '06, ACM, pp. 336–343.

[342] TUDDENHAM, P., KIRK, D., AND IZADI, S. Graspables revisited: Multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2010), CHI '10, ACM, pp. 2223–2232.

[343] TUDDENHAM, P., AND ROBINSON, P. Territorial coordination and workspace awareness in remote tabletop collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2009), CHI '09, ACM, pp. 2139–2148.

[344] URACOLI. µracoli - The µController Radio Communication Library. `http://www.nongnu.org/uracoli/`, 2016 (accessed March 5, 2016).

[345] VAN RHIJN, A., AND MULDER, J. D. Spatial input device structure and bimanual object manipulation in virtual environments. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2006), VRST '06, ACM, pp. 51–60.

[346] VASUDEVAN, R., LOBATON, E., KURILLO, G., BAJCSY, R., BERNARDIN, T., HAMANN, B., AND NAHRSTEDT, K. A methodology for remote virtual interaction in teleimmersive environments. In *Proc. of Multimedia Systems 2010* (New York, NY, USA, 2010), ACM Press, pp. 281–292.

[347] VEIT, M., CAPOBIANCO, A., AND BECHMANN, D. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2009), VRST '09, ACM, pp. 51–58.

[348] VEIT, M., CAPOBIANCO, A., AND BECHMANN, D. Dynamic decomposition and integration of degrees of freedom for 3-d positioning. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology* (New York, NY, USA, 2010), VRST '10, ACM, pp. 131–134.

[349] VERTEGAAL, R., WEEVERS, I., SOHN, C., AND CHEUNG, C. Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction. In *Proc. of CHI 2003* (New York, NY, USA, 2003), ACM Press, pp. 521–528.

[350] VOGEL, D., AND BALAKRISHNAN, R. 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* (New York, NY, USA, 2005), UIST '05, ACM, pp. 33–42.

[351] VRAC. Virtual reality application center, Iowa State University. `http://www.vrac.iastate.edu/c6.php`, 2016 (accessed March 5, 2016).

[352] WAGNER, J., HUOT, S., AND MACKAY, W. Bitouch and bipad: Designing bimanual interaction for hand-held tablets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2012), CHI '12, ACM, pp. 2317–2326.

[353] WALLACE, J. R., SCOTT, S. D., STUTZ, T., ENNS, T., AND INKPEN, K. Investigating teamwork and taskwork in single- and multi-display groupware systems. *Personal Ubiquitous Comput. 13*, 8 (Nov. 2009), 569–581.

[354] WANG, F., CAO, X., REN, X., AND IRANI, P. Detecting and leveraging finger orientation for interaction with direct-touch surfaces. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology* (2009), ACM, pp. 23–32.

[355] WANG, Y., MACKENZIE, C. L., AND SUMMERS, V. A. Object manipulation in virtual environments: Human bias, consistency and individual differences. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 1997), CHI EA '97, ACM, pp. 349–350.

[356] WANG, Y., MACKENZIE, C. L., SUMMERS, V. A., AND BOOTH, K. S. The structure of object transportation and orientation in human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1998), CHI '98, ACM Press/Addison-Wesley Publishing Co., pp. 312–319.

[357] WARE, C. Using hand position for virtual object placement. *Vis. Comput. 6*, 5 (Nov. 1990), 245–253.

[358] WARE, C., AND ARSENAULT, R. Frames of reference in virtual object rotation. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization* (New York, NY, USA, 2004), APGV '04, ACM, pp. 135–141.

[359] WARE, C., AND BALAKRISHNAN, R. Reaching for objects in vr displays: Lag and frame rate. *ACM Trans. Comput.-Hum. Interact. 1*, 4 (Dec. 1994), 331–356.

[360] WARE, C., AND ROSE, J. Rotating virtual objects with real handles. *ACM Trans. Comput.-Hum. Interact. 6*, 2 (June 1999), 162–180.

[361] WEINTRAUB, D. J. Rectangle discriminability: Perceptual relativity and the law of pragnanz. *Journal of Experimental Psychology 88*, 1 (1971), 1.

[362] WEISS, M., JENNINGS, R., KHOSHABEH, R., BORCHERS, J., WAGNER, J., JANSEN, Y., AND HOLLAN, J. D. Slap widgets: Bridging the gap between virtual and physical controls on tabletops. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2009), CHI EA '09, ACM, pp. 3229–3234.

[363] WIDESTRÖM, J., AXELSSON, A.-S., SCHROEDER, R., NILSSON, A., HELDAL, I., AND ABELIN, Å. The collaborative cube puzzle: A comparison of virtual

and real environments. In *Proceedings of the Third International Conference on Collaborative Virtual Environments* (New York, NY, USA, 2000), CVE '00, ACM, pp. 165–171.

[364] WIESENDANGER, M., KAZENNIKOV, O., PERRIG, S., AND KALUZNY, P. *Two hands – one action: The problem of bimanual coordination.* Academic Press, San Diego, CA, US, 1996, pp. 283–300.

[365] WIGDOR, D., JIANG, H., FORLINES, C., BORKIN, M., AND SHEN, C. Wespace: The design development and deployment of a walk-up and share multi-surface visual collaboration system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2009), CHI '09, ACM, pp. 1237–1246.

[366] WOBBROCK, J. O., CUTRELL, E., HARADA, S., AND MACKENZIE, I. S. An error model for pointing based on fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, ACM, pp. 1613–1622.

[367] WOBBROCK, J. O., JANSEN, A., AND SHINOHARA, K. Modeling and predicting pointing errors in two dimensions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2011), CHI '11, ACM, pp. 1653–1656.

[368] WOLFF, R., ROBERTS, D. J., STEED, A., AND OTTO, O. A review of telecollaboration technologies with respect to closely coupled collaboration. *Int. J. Comput. Appl. Technol. 29* (June 2007), 11–26.

[369] WU, M., AND BALAKRISHNAN, R. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proc. UIST 2003* (2003), ACM Press, pp. 193–202.

[370] WYSS, H. P., BLACH, R., AND BUES, M. isith-intersection-based spatial interaction for two hands. In *3D User Interfaces, 2006. 3DUI 2006. IEEE Symposium on* (2006), IEEE, pp. 59–61.

[371] YAMANISHI, J.-I., KAWATO, M., AND SUZUKI, R. Two coupled oscillators as a model for the coordinated finger tapping by both hands. *Biological Cybernetics 37*, 4 (1980), 219–225.

[372] YEE, K.-P. Two-handed interaction on a tablet display. In *Ext. Abstracts CHI 2004* (2004), ACM Press, pp. 1493–1496.

[373] ZABRAMSKI, S. Careless touch: A comparative evaluation of mouse, pen, and touch input in shape tracing task. In *Proceedings of the 23rd Australian Computer-Human Interaction Conference* (New York, NY, USA, 2011), OzCHI '11, ACM, pp. 329–332.

[374] ZELEZNIK, R., BRAGDON, A., ADEPUTRA, F., AND KO, H.-S. Hands-on math: A page-based multi-touch and pen desktop for technical work and problem solving. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2010), UIST '10, ACM, pp. 17–26.

[375] ZELEZNIK, R., FORSBERG, A., AND STRAUSS, P. Two Pointer Input for 3D Interaction. In *Proc. of SI3D '97* (1997), pp. 115–ff.

[376] ZHAI, S. *Human Performance in Six Degree of Freedom Input Control, Ph.D. Thesis.* Univ. of Toronto, 1995.

[377] ZHAI, S. Characterizing computer input with fitts' law parameters: The information and non-information aspects of pointing. *Int. J. Hum.-Comput. Stud. 61*, 6 (Dec. 2004), 791–809.

[378] ZHAI, S., AND MILGRAM, P. Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1998), CHI '98, ACM Press/Addison-Wesley Publishing Co., pp. 320–327.

[379] ZHAI, S., MILGRAM, P., AND BUXTON, W. The influence of muscle groups on performance of multiple degree-of-freedom input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1996), CHI '96, ACM, pp. 308–315.

[380] ZHAI, S., MORIMOTO, C., AND IHDE, S. Manual and gaze input cascaded (magic) pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 1999), CHI '99, ACM, pp. 246–253.

[381] ZHAI, S., AND SENDERS, J. W. Investigating coordination in multidegree of freedom. In *In 41st Annual Meeting of Human Factors and Ergonomics Society* (1997), Society, pp. 1249–1253.

[382] ZHANG, Z., ZHANG, F., CHEN, H., LIU, J., WANG, H., AND DAI, G. Left and right hand distinction for multi-touch tabletop interactions. In *Proceedings of the 19th International Conference on Intelligent User Interfaces* (New York, NY, USA, 2014), IUI '14, ACM, pp. 47–56.

[383] ZILLNER, J., RHEMANN, C., IZADI, S., AND HALLER, M. 3d-board: A whole-body remote collaborative whiteboard. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 2014), UIST '14, ACM, pp. 471–479.

[384] ZIMMERMAN, J., FORLIZZI, J., AND EVENSON, S. Research through design as a method for interaction design research in hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2007), CHI '07, ACM, pp. 493–502.

Typeset November 19, 2016