



**IEEE Press Series on Human-Machine Systems**  
Giancarlo Fortino, Series Editor

# Computer-Supported Collaboration

THEORY AND PRACTICE

WEIDONG HUANG, MARK BILLINGHURST,  
LEILA ALEM, CHUN XIAO, TROELS RASMUSSEN



 **IEEE Press**

**WILEY**



## **Computer-Supported Collaboration**

**IEEE Press**  
445 Hoes Lane  
Piscataway, NJ 08854

**IEEE Press Editorial Board**  
Sarah Spurgeon, *Editor-in-Chief*

Moeness Amin  
Jón Atli Benediktsson  
Adam Drobot  
James Duncan

Ekram Hossain  
Brian Johnson  
Hai Li  
James Lyke  
Joydeep Mitra

Desineni Subbaram Naidu  
Tony Q. S. Quek  
Behzad Razavi  
Thomas Robertazzi  
Diomidis Spinellis

# Computer-Supported Collaboration

Theory and Practice

*Weidong Huang*

University of Technology Sydney  
Sydney, Australia

*Mark Billingham*

University of Auckland  
Auckland, New Zealand

and

University of South Australia  
Adelaide, Australia

*Leila Alem*

University of Technology Sydney  
Sydney, Australia

*Chun Xiao*

University of Technology Sydney  
Sydney, Australia

*Troels Rasmussen*

Aarhus University  
Aarhus, Denmark

 **IEEE Press**

**WILEY**

Copyright © 2024 by The Institute of Electrical and Electronics Engineers, Inc.  
All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.  
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at [www.copyright.com](http://www.copyright.com). Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permission>.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at [www.wiley.com](http://www.wiley.com).

***Library of Congress Cataloging-in-Publication Data applied for:***

Hardback: 9781119719762

Cover Design: Wiley

Cover Image: © XH4D/Getty Images

Set in 9.5/12.5pt STIXTwoText by Straive, Chennai, India

## Contents

**About the Authors** *xiii*

**Acknowledgments** *xv*

<b>1</b>	<b>Remote Collaboration on Physical Tasks</b>	<b>1</b>
1.1	Introduction	1
1.2	Remote Collaboration in Perspective	2
1.3	Book Audience	5
	References	6
<b>2</b>	<b>Communication Models for Remote Guidance</b>	<b>7</b>
2.1	Introduction	7
2.2	Overview of Communication Models	9
2.2.1	Linear Communication Models	12
2.2.2	Nonlinear Communication Models	13
2.2.3	Summary	18
2.3	Applying Communication Models	19
2.4	Communication Behaviors in AR Conferencing	22
2.5	A Communication Model for AR	32
2.6	Conclusions	34
	References	35
<b>3</b>	<b>Communication Cues in Augmented Remote Collaboration</b>	<b>41</b>
3.1	Introduction	41
3.2	The Research Landscape – Trends Over Time	43
3.3	Communication in Augmented Remote Collaboration	48
3.3.1	Context for Communication	48
3.3.2	Verbal Communication Cues	50

- 3.3.3 Visual Communication Cues 51
  - 3.3.3.1 Annotations 51
  - 3.3.3.2 Pointers 52
  - 3.3.3.3 Body Gestures 54
  - 3.3.3.4 Eye Gaze 59
  - 3.3.3.5 Virtual Replica of Objects 60
- 3.3.4 Haptic Communication Cues 61
- 3.3.5 Empathic Communication Cues 62
  - 3.3.5.1 Facial Expression 63
  - 3.3.5.2 Eye Gaze and Body Gestures 63
  - 3.3.5.3 Paralinguistic Speech Signals 64
  - 3.3.5.4 Physiological Cues 64
- 3.4 Challenges 65
  - 3.4.1 Multimodality Communication Modeling 65
  - 3.4.2 Ease of Use: Embrace New Technologies 65
- 3.5 Future Directions 67
  - 3.5.1 Natural Communication-Based Remote Collaboration 67
  - 3.5.2 Cloud-Based Remote Collaboration 67
  - 3.5.3 Empathic Remote Collaboration 68
- 3.6 Conclusion 68
  - References 69
  
- 4 Communication Cues for Remote Guidance 81**
  - 4.1 Introduction 81
  - 4.2 Explicit Communication Cues 84
    - 4.2.1 Pointer 85
      - 4.2.1.1 Cursor 85
      - 4.2.1.2 Laser Pointer 85
      - 4.2.1.3 Stick Mounted on a Robot 86
      - 4.2.1.4 Eye-Gaze pointer and Head Pointer 86
      - 4.2.1.5 Tracked Marker Annotations 87
    - 4.2.2 Sketches and Annotations 88
      - 4.2.2.1 Draw on Video 89
      - 4.2.2.2 Draw into Workspace 89
      - 4.2.2.3 Tracked Drawing Annotations 90
      - 4.2.2.4 Combination 92
    - 4.2.3 Hand Gesture 93
      - 4.2.3.1 Video Feed of Hand Gestures 93
      - 4.2.3.2 Superimposed Hand Gesture 94
      - 4.2.3.3 Combination 96
    - 4.2.4 Object Models 97
    - 4.2.5 Comparison of Explicit Communication Cues 98



4.3	Implicit Communication Cues	100
4.4	Challenges and Future Directions	104
4.4.1	Technical and Technological Limitations	104
4.4.2	Empathic, Mental Effort, and Fatigue Cues	104
4.4.3	Evaluation Metrics and Methodologies	105
4.5	Conclusion	105
	Acknowledgment	106
	References	107
<b>5</b>	<b>Communicating Eye Gaze Cues in Remote Collaboration on Physical Tasks</b>	<b>115</b>
5.1	Introduction	115
5.2	The Changing Research Landscape – Research Topic Trends and Teams over the Past Two Decades	117
5.2.1	Method of Data Collection from Scopus	117
5.2.2	Bibliometric Dataset Cleaning for Topic and Coauthorship Network Visualization	117
5.2.3	An Overview of Research Topic Changes and Collaborations	118
5.2.4	Screening for Detailed Article Review	121
5.3	Categorization of System Setup Based on the Screened Publications	122
5.3.1	Eye Gaze Overlaid on a Fixed Screen or Head-Mounted Display	125
5.3.2	Augmented Eye Gaze Projected on a Fixed Workspace	126
5.3.3	Augmented Eye Gaze in Virtual Environment	127
5.4	Gaze Visualization	127
5.4.1	Gaze Cursor in Two-Dimensional View	127
5.4.2	Gaze Visible as a Projected Pointer in Physical Workspace	128
5.4.3	Gaze as a Cursor or Ray-Cast in AR/VR/MR Carried by Avatars	128
5.4.4	Gaze Visualization with Controlled Visibility	129
5.4.5	Effectiveness of Gaze Visibility and Visualization	129
5.5	Functionality of Tracked Gaze in Remote Guidance on Physical Tasks	130
5.6	Challenges of Utilizing Eye Tracking in Remote Collaboration	132
5.7	Future Directions	133
5.8	Conclusion	135
	References	135
<b>6</b>	<b>Evaluating Augmented Reality Remote Guidance Systems</b>	<b>143</b>
6.1	Introduction	143
6.2	Evaluation Methods for Collaborative AR	148

- 6.3 Case Studies From Example Systems 152
- 6.3.1 Impact of Gaze Tracking on Remote Guidance Collaboration 153
- 6.3.2 Sharing Awareness Cues in Collaborative Mixed Reality 158
- 6.4 Guidelines 163
- 6.5 Directions for Research 164
- 6.6 Conclusion 166
- References 166

**7 Supporting Remote Hand Gestures over the Workspace Video 173**

- 7.1 Introduction 173
- 7.2 Related Work 176
- 7.2.1 Systems for Remote Guidance 176
- 7.2.2 Supporting Remote Gestures 176
- 7.3 HandsOnVideo 177
- 7.3.1 Worker Interface Design 178
- 7.3.2 Helper Interface Design 179
- 7.3.3 The Platform and Technical Specifications 181
- 7.4 User Testing 183
- 7.4.1 Design 183
- 7.4.2 Participants 184
- 7.4.3 Tasks 184
- 7.4.4 Procedure 185
- 7.4.5 Observations 185
- 7.4.6 Questionnaire Results 187
- 7.5 Discussion 189
- 7.6 Conclusion and Future work 190
- Acknowledgment 191
- References 191

**8 Gesturing in the Air in Supporting Full Mobility 195**

- 8.1 Introduction 195
- 8.2 Background 197
- 8.2.1 Shared Visual Space 197
- 8.2.2 Remote Gestures 197
- 8.2.3 Supporting Remote Gestures 198
- 8.2.3.1 Agent-Based Remote Gestures 198
- 8.2.3.2 Digital Annotations/Gestures 199
- 8.2.3.3 Projected Hands 199
- 8.2.3.4 Hands over Workspace Videos 199

8.3	System Overview	200
8.3.1	Hands in the Air	200
8.3.2	Hardware and Software Implementation	200
8.3.3	How the System Works	201
8.4	Usability Study	203
8.4.1	Design	203
8.4.2	Participants	204
8.4.3	Task	204
8.4.4	Procedure	205
8.4.5	Results	205
8.4.5.1	Observations	205
8.4.5.2	Usability Ratings	206
8.4.5.3	User Comments	208
8.5	Discussion	209
8.5.1	Usability	209
8.5.2	Gesturing in the Air	210
8.5.3	Limitation of the Studies	210
8.6	Concluding Remarks and Future Work	210
	Acknowledgment	211
	References	211
<b>9</b>	<b>Sharing Hand Gesture and Sketch Cues with a Touch User Interface</b>	<b>215</b>
9.1	Introduction	215
9.2	Related Work	217
9.2.1	Pointer Cues	217
9.2.2	Sketch Cues	218
9.2.3	Hand Gesture Cues	219
9.2.4	Discussion	219
9.3	Methods and Materials	220
9.3.1	Prototype System	220
9.3.1.1	Hardware	221
9.3.1.2	The User Interfaces and How HandsInTouch Works	222
9.3.1.3	Hand Gestures and Sketches	224
9.3.2	User Study Design	225
9.3.2.1	Tasks	225
9.3.2.2	Measurement	226
9.3.2.3	Participants	227
9.3.2.4	Procedure	228
9.4	Results	228
9.4.1	Task Performance and Task Load	228

9.4.2	Overall Usability	230
9.4.3	Sketch Usefulness	230
9.4.4	Observations and Log Data Results	231
9.4.5	Qualitative Results	233
9.5	Discussion	234
9.5.1	Discussion of Our Results	234
9.5.2	Implications for the Design of Remote Guidance Systems	236
9.5.3	Implication for the Application of the System	236
9.6	Limitations	237
9.7	Conclusion	238
	Acknowledgment	238
	References	239
<b>10</b>	<b>Augmenting Hand Gestures in 3D Mixed Reality</b>	<b>243</b>
10.1	Introduction	243
10.2	Related Work	245
10.2.1	Remote Guidance Systems	245
10.2.2	Collaborative Virtual Environments	247
10.2.3	Implicit Visual Communication Cues and Cultural Factors	248
10.3	System Overview	248
10.3.1	Technical Challenges	248
10.3.2	Hardware, Software and Implementation	250
10.3.3	Overall Architecture	251
10.4	Evaluation	253
10.4.1	Method	253
10.4.2	Procedure	254
10.4.3	Results and Discussion	255
10.4.3.1	Observations	255
10.4.3.2	Usability Ratings	255
10.4.3.3	User Experiences	259
10.5	A Comparison of User Ratings between HandsInAir and HandsIn3D	260
10.6	Conclusion and Future Work	262
	Acknowledgment	264
	References	265
<b>11</b>	<b>Supporting Tailorability to Meet Individual Task Needs</b>	<b>269</b>
11.1	Introduction	269
11.2	Component-Based Design of RemoteAssistKit	271
11.2.1	Helper App	272
11.2.2	Worker App	275

11.2.3	Supplementary Worker App	275
11.2.4	Tailoring Mounting Equipment	276
11.3	Identifying Tailorable Aspects of Remote Assistance	276
11.4	How Users Tailor Remote Assistance	281
11.4.1	Users Tailored Guidance Format and Task Space Capturing to the Requirements of Scenarios	283
11.4.2	Awareness of the Collaborating Partner's Composition of Components Is Important	286
11.5	The Importance of Nonverbal Guidance Depends on the Knowledge Relationship	287
11.6	Sharing of Machine Sounds Is Important for Remote Troubleshooting	288
11.7	High-Resolution Views Are Important for Remote Product Quality Optimization	289
11.8	The Manufacturing Context Poses a Challenge for Creating 3D Reconstructions with Depth Cameras	290
11.9	Multiple Cameras Support Workspace Awareness in Large Industrial Task Spaces	290
11.10	Concluding Remarks	293
	References	294
<b>12</b>	<b>Supporting Workspace Awareness with Augmented Reality-Based Multi-camera Visualization and Tracking</b>	<b>299</b>
12.1	Introduction	299
12.2	Augmented Reality for Supporting Awareness During Multi-camera Remote Assistance	302
12.2.1	Core Functionality of SceneCam and CueCam	303
12.2.2	Supporting the Helper's Awareness with SceneCam	305
12.2.2.1	Ad Hoc Creation of (Scene Camera, Work Area)-Pairs	305
12.2.2.2	Focus-in-Context Views	306
12.2.2.3	Automatic View Selection	309
12.2.3	Supporting the Workers Awareness with CueCam	309
12.3	Future Research on Multi-camera Remote Assistance	314
12.3.1	Improving Focus-in-Context Views	314
12.3.1.1	Supporting Accurate Distal Pointing in Exocentric Focus-in-Context View	314
12.3.1.2	Combining Scene Cameras and Head-Mounted Camera	315
12.3.1.3	Modeling the Environment in Exocentric Focus-in-Context View	316
12.3.2	Improved Automatic View Selection Algorithm	316
12.3.3	Automatic Camera Configuration Detection for Adaptive AR Multi-camera Remote Assistance	318

- 12.3.4 Improving Color Cue to Ease Mapping of Helper’s Location 318
- 12.3.5 Study of Multi-camera Remote Assistance in the Manufacturing Industry 320
- 12.4 Discussion of 2D vs. 3D Workspace Information 320
- 12.5 Concluding Remarks 321
- References 322

**13 Industrial Applications, Current Challenges, and Future Directions 327**

- 13.1 Introduction 327
- 13.2 Remote Guidance Systems 327
  - 13.2.1 Alfa Laval Remote Guidance 328
  - 13.2.2 TEXO Remote Guidance Software 328
  - 13.2.3 Kiber 3 329
  - 13.2.4 Shipznet RG 300 – Remote Guidance Service Connectivity 329
  - 13.2.5 HPE Visual Remote Guidance Service 329
  - 13.2.6 Dynamics 365 Remote Assist 330
  - 13.2.7 ABB Remote Assistance 330
  - 13.2.8 CSIRO ReMoTe 330
  - 13.2.9 Summary 331
- 13.3 Technical, Ethical, and Research Challenges and Future Directions 331
  - 13.3.1 Ergonomically Tested Devices and Privacy and Ethical Aspects of Remote Guidance 331
  - 13.3.2 Network Connection and Information Delay 332
  - 13.3.3 Reproducing the Environment of Face-to-Face Collaboration for Remote Collaboration 335
  - 13.3.4 Replacing a Communication Cue with Another Cue of a Different Modality 336
- 13.4 Conclusion 338
- References 339

**Index 343**

## About the Authors

**Weidong Huang** received his PhD in computer science from the University of Sydney. He is currently Associate Professor at the University of Technology Sydney. He also has formal training in experimental psychology and professional experience in psychometrics. He is the author of over 150 publications. His research interests are in human–computer interaction, visualization, and data science. He is a founding chair of the Technical Committee on Visual Analytics and Communication for IEEE SMC Society and a guest editor of a number of SCI indexed journals. He has served as a conference chair, a program committee chair, or an organization chair for a number of international conferences and workshops.

**Mark Billingham** is currently Director of the Empathic Computing Laboratory and Professor at the University of South Australia in Adelaide, Australia, and also at the University of Auckland in Auckland, New Zealand. He earned a PhD in 2002 from the University of Washington and conducts research on how virtual and real worlds can be merged, publishing over 650 papers on augmented reality, virtual reality, remote collaboration, empathic computing, and related topics. In 2013, he was elected as a Fellow of the Royal Society of New Zealand, and in 2019 he was given the ISMAR Career Impact Award in recognition for lifetime contribution to AR research and commercialization. In 2022, he was selected for the ACM SIGCHI Academy, for leading human–computer interaction researchers, and also selected for the IEEE VGTC VR Academy for leading VR researchers.

**Leila Alem** is an adjunct professor at the UTS faculty of engineering and information technology. She was a principal research scientist at the CSIRO Digital Productivity and Services based in Sydney. Her formal training is in artificial intelligence and cognitive psychology. Over the past 23+ years of her career at CSIRO, she has designed, evaluated, and delivered ICT solutions and services for industries such as aviation, mining, manufacturing, and health. Leila has an established international profile in human–computer interaction with a special interest in

enhancing user experiences through novel ways of working together using mobile and wearable technologies. She is the editor of two books with Springer on mobile collaborative augmented reality systems. Her main focus of research has been in the area of human factors in computer mediated collaboration settings. Drawing on cognitive psychology, social science and human–computer interface research she has investigated the media factors, the cognitive factors, and the social factors at play in telepresence systems and environments.

**Chun Xiao** received her PhD in computer science from Otto-von-Guericke University Magdeburg. She was a postdoctoral researcher at the University of Technology Sydney, conducting research on remote collaboration. She has published papers in human–computer interaction and visualization.

**Troels Rasmussen** received his PhD in computer science from Aarhus University Denmark. His research interests are in augmented reality and human–computer interaction. He was a postdoctoral researcher at the Aarhus University.



## Acknowledgments

Our research on remote collaboration on physical tasks started from an R&D project with an industry partner in the mining industry more than 15 years ago. The project was to support the changing role of mining operators in the face of mining automation. The digitally enabled collaborations that take place in these remote operations involved an operator collocated with the object of the collaboration, which is a physical object (a piece of mining equipment) and a remote expert. Our initial research has drawn on previous work on remote collaboration on physical tasks by Susan Fussel and Mark Billinghurst. Since then, we have collaborated with researchers, students, and a number of organizations including one start-up and a commercial partner. They have contributed to the research results that are presented in this book directly or indirectly in various ways, and we express our sincere gratitude to them. Without their contribution, participation, and support, this book would not have come to fruition. In particular, we thank our coauthors: Franco Tecchia, Seungwon Kim, Mathew Wakefield, and Henry Been-Lirn Duh, for giving us permissions to include the coauthored published works in this book. For the same, we thank the publishers as well: Springer, J.UCS Consortium, Bentham Open, and Elsevier. We also thank Tiare Feuchtner and Kaj Grønbaek for their contribution to the work discussed in Chapters 11 and 12. Finally, we thank Boeing R&D for the opportunity to conduct a trial within their operation in Seattle, which has led to the successful commercialisation of our research effort.

We extend our sincere thanks and appreciation to the editors Indirakumari Siva, Ranjith Kumar Thanigasalam, Teresa Netzler, and Victoria Bradshaw and the production team from John Wiley & Sons for their professional support throughout this project.



# 1

## Remote Collaboration on Physical Tasks

### 1.1 Introduction

Remote collaboration on physical tasks (or remote guidance/remote assistance) typically involves one or more remote helpers guiding one or more local workers to work collaboratively on the manipulation of physical objects [1, 2]. In this type of remote collaboration, both the workers and helpers are physically distributed. On the one hand, the workers have direct access to the physical objects to be worked on but do not have full skills or knowledge on how to operate or manipulate them; thus, they need to receive help from the remote helpers. On the other hand, the remote helpers know how, but do not have physical access to the objects [3]. Technologies that support remote guidance can greatly improve the productivity and safety of tasks by allowing experts to provide timing guidance and training to individuals remotely without having to travel on-site. It has a wide range of applications in industrial domains (e.g. [4]) and has the potential to revolutionize those industries in terms of how the business operates and how service can be provided to their customers, from manufacturing and construction to healthcare and education, to name a few.

With recent advances in networking, augmented reality (AR), virtual reality (VR), mobile and wearable technologies, it has become increasingly possible in practice to enable helpers to remotely guide individuals in performing complex physical tasks with precision and efficiency [5]. Given the increasing demand for remote guidance technologies from industries and increasing interest and effort in research from academics, this research book explores the latest and typical developments in remote guidance technologies and provides comprehensive reviews of the current state-of-the-art research in this field, including our own research findings and developments in the past 15 years.

## 1.2 Remote Collaboration in Perspective

The rest of the book has 12 chapters, each focusing on a specific aspect of remote collaboration research. Both technology and communication are essential elements of remote collaboration, and understanding whether and how technology impacts communication behaviors is important for the design of remote collaboration systems. However, this is an area that has not been well-researched. The technology impact can be predicted using communication models. Thus, we dedicate the next chapter of our book to the discussion of how existing communication models can be used to predict the impact of different AR technologies in remote collaboration and if a new communication model needs to be developed. More specifically, we provide a review of various existing communication models and show how they can be used to analyze communication in both AR and non-AR interfaces for remote guidance on physical tasks. We also discuss the limitations of current models, identify research gaps, and explore possible further developments.

The third chapter provides a review of communication cues in remote collaboration. It starts with an overview of the research landscape over the past three decades and then investigates the communication context based on which a remote collaboration is conducted. We categorize communication cues in remote collaboration systems as verbal, visual, haptic, and empathic communication cues and review the systems and experiments that studied each of them to identify advantages and limitations under different situations. Finally, we summarize and address the challenges of multimodality communication modeling and system design for high usability and suggest potential future research directions for augmented remote collaboration system design aiming at effectiveness, reliability, and ease of use.

For remote guidance on physical tasks, in addition to verbal communications, how to convey other communication cues effectively has been researched extensively. Given the importance and variety of possible communication cues as outlined in the third chapter, we presented a review in the fourth chapter to summarize the communication cues being used, approaches that implement the cues, and their effects on remote guidance on physical tasks [6]. In this chapter, we categorize the communication cues into explicit and implicit ones and report our findings. Our review indicates that a number of communication cues have been shown to be effective in improving system usability and helping collaborators to achieve optimal user experience and task performance. More specifically, there is a growing interest in providing a combination of multiple explicit communication cues to cater for the needs of different task purposes and in providing combination of explicit and implicit communication cues.

Although technology for remote collaboration is becoming increasingly more essential and affordable, and eye gaze is an important cue for human–human communication, there is much that remains to be done to explore the use of gaze in remote collaboration, especially for collaboration on physical tasks. Recent advancement in eye tracking technologies enables gaze input to be added to collaborative systems, especially for remote guidance and is expected to bring more promising opportunities to reduce misunderstanding and improve effectiveness. The fifth chapter surveys publications with respect to eye tracking-supported collaborative physical work under remote guidance. We categorize the prototypes and systems presented according to four metrics ranging from eye-tracked subjects to gaze visualization. Then, we summarize the experimental and investigation findings to have an overview of the eye tracking mechanism in remote physical collaboration systems, as well as the roles that eye gaze and its visualization play in common understanding, referential, and social copresence practices.

The sixth chapter provides a summary of how to conduct evaluation studies of AR-based remote guidance systems. As previously discussed in this book, communication is an essential part of remote collaboration, and many technologies have been developed to enable people to better connect and communicate with one another. However, the impact of these technologies can only be measured through conducting evaluation studies and measuring how the technologies change communication behavior between real people. Therefore, the purpose of this chapter is to help the readers become more proficient in their own evaluation studies and create research outputs that will inspire others in the field. More specifically, in this chapter, we present evaluation case studies, derive a number of design guidelines, and discuss methods that can be used to create robust evaluation studies. Finally, this chapter concludes with a list of possible research directions.

From the seventh chapter, we introduce a range of typical remote guidance systems. These systems were developed with different configurations to meet different collaboration requirements and to serve as platforms for us to investigate specific research questions. First, in this chapter, we present a remote guidance system called HandsOnVideo [7], a system that uses a near-eye display to support mobility and unmediated representations of hands to support remote gestures, enabling a remote helper guiding a mobile worker working in nontraditional-desktop environments. The system was designed and developed using a participatory design approach, which allowed us to test and trial a number of design ideas. It also enabled us to understand from a user’s perspective some of the design tradeoffs. The usability study with end users indicated that the system is useful and effective. The users were also positive about using the near-eye display for mobility and instructions and using unmediated representations of hands for remote gestures.

The eighth chapter introduces HandsInAir [8], a wearable system for remote guidance. This system is designed to support the mobility of the collaborators

and provide easy access to remote expertise. HandsInAir draws on the richness of hand gestures for remote guiding and implements a novel approach that supports unmediated remote gestures and allows the helper to perform natural gestures by hands without the need for physical support. A usability study was also conducted demonstrating the usefulness and usability of HandsInAir. More specifically, the participants were positive about the mobility support provided by the system to the collaborators. According to their feedback, the mobility support allows workers to access a remote helper more easily. Also, helpers are enabled to continuously engage with the system and their partner when they move around during the guiding process. Participants who played the role of helper also considered gesturing in the air as being intuitive and effective.

The ninth chapter introduces HandsInTouch [9], which supports a unique remote collaboration gesture interface by including both raw hand gestures and sketch cues on a live video or still images. We also conducted a user study comparing remote collaboration with the interface that combines hand gestures and sketching (the HandsInTouch interface) to one that only used hand gestures when solving two tasks: Lego assembly and repairing a laptop. It was found from the study that adding sketch cues improved the task completion time, only with the repairing task, as this had complex object manipulation, and that using gesture and sketching together created a higher task load for the user. The implications of our findings for system design and application are also discussed in the chapter.

The tenth chapter describes Handsin3D [10], a system that uses three-dimensional (3D) real-time capturing and rendering of both the remote workspace and the helper's hands and creates a 3D shared visual space as a result of colocating the remote workspace with the helper's hands. The 3D shared space is displayed on a head-tracked stereoscopic hand-mounted display (HMD) that allows the helper to perceive the remote space in 3D as well as guide in 3D. A user study conducted with the system reveals that the unique feature of HandsIn3D is the integration of the projection of the helper's hands into the 3D workspace of the worker. Not only does this integration gives users flexibility in performing more natural hand gestures and ability in perceiving spatial relationship of objects more accurately but also offers greater sense of copresence and interaction.

The eleventh chapter introduces a component-based tailorable remote assistance system called RAK. The design and development of RAK were informed by the results and findings of an interview study with employees of a manufacturing industry. Then, an experimental simulation with RAK that was conducted at a technical college for plastic manufacturing was briefly described. A large part of the chapter was devoted to our discussion and reflection on the results and observations of the user studies. It is encouraging that we are able to derive some

meaningful and unexpected new insights, which could guide the directions of future work. These include the tailoring behaviors of both workers and helpers, sharing machine sound from the workspace to the helper, and supporting workspace awareness with multi-camera setups.

The twelfth chapter introduces two multi-camera AR research prototypes, SceneCam and CueCam. These two systems are developed to help collaborators maintain awareness of each other in large workspaces. Multi-camera remote assistance has some benefits over using one camera from the point of view of the worker, most notably the view independence of the helper. However, in this chapter, we point out the challenges that stand in the way of obtaining good workspace awareness when using multiple cameras and demonstrate with the two systems how AR visualization and tracking can be used to address these awareness challenges in various ways.

The final chapter introduces some industrial systems that support remote guidance on physical tasks. Each of these industrial systems was designed to meet specific design and/or business purposes. Current challenges and possible future directions are also discussed. These include ergonomically tested devices and privacy and ethical aspects of remote guidance, network connection, and information delay, reproducing the environment of face-to-face collaboration for remote collaboration, and replacing a communication cue with another cue of a different modality. Apart from these topics, the chapter concludes the book with other possible directions being mentioned, including artificial intelligence and cloud-based remote guidance support, embedment, and integration of cognitive, physiological, empathic, and multimodal communication cues, investigation of possible effects of human factors, language, social and cultural factors, and more rigors and empirically validated evaluation frameworks, design principles, metrics, and methodologies for remote collaboration on physical tasks.

### **1.3 Book Audience**

This book is for researchers, engineers, scientists, and practitioners who are interested in the research of remote collaboration and its potential applications in various industrial domains. Academics and postgraduate students in science and engineering will also find this book useful as a comprehensive reference book. It provides a comprehensive overview of and detailed insights into the current state-of-the-art research and the potential future directions for the topic. We hope that this book will inspire new research and innovation, and ultimately lead to new theories and development of more effective and efficient remote collaboration systems and tools to meet real-world needs.

## References

- 1 Kraut, R.E., Fussell, S.R., and Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction* 18 (1-2): 13–49. [https://doi.org/10.1207/S15327051HCI1812\\_2](https://doi.org/10.1207/S15327051HCI1812_2).
- 2 Gergle, D., Kraut, R.E., and Fussell, S.R. (2013). Using visual information for grounding and awareness in collaborative tasks. *Human-Computer Interaction* 28 (1): 1–39. <https://doi.org/10.1080/07370024.2012.678246>.
- 3 Kim, S., Billinghamurst, M., and Kim, K. (2020). Multimodal interfaces and communication cues for remote collaboration. *Journal on Multimodal User Interfaces* 14: 313–319. <https://doi.org/10.1007/s12193-020-00346-8>.
- 4 Kiber. Kiber 3. <https://kiber.tech/company/> (accessed 3 July 2022).
- 5 Wang, P., Bai, X., Billinghamurst, M. et al. (2021). AR/MR remote collaboration on physical tasks: a review. *Robotics and Computer-Integrated Manufacturing* 72 (C): <https://doi.org/10.1016/j.rcim.2020.102071>.
- 6 Huang, W., Wakefield, M., Rasmussen, T.A. et al. (2022). A review on communication cues for augmented reality based remote guidance. *Journal on Multimodal User Interfaces* 16: 239–256. <https://doi.org/10.1007/s12193-022-00387-1>.
- 7 Alem, L., Huang, W., and Tecchia, F. (2011). Supporting the changing roles of maintenance operators in mining: a human factors perspective. *The Ergonomics Open Journal*. 4: 81–92. <https://doi.org/10.2174/1875934301104010081>.
- 8 Huang, W. and Alem, L. (2013). Gesturing in the air: supporting full mobility in remote collaboration on physical tasks. *Journal of Universal Computer Science*. 19: 1158–1174. <https://doi.org/10.3217/jucs-019-08-1158>.
- 9 Huang, W., Kim, S., Billinghamurst, M., Alem, L. (2019) Sharing hand gesture and sketch cues in remote collaboration. *Journal of Visual Communication and Image Representation*. 58, 428–438. <https://doi.org/https://doi.org/10.1016/j.jvcir.2018.12.010>.
- 10 Huang, W., Alem, L., Tecchia, F., and Duh, H. (2018). Augmented 3D hands: a gesture-based mixed reality system for distributed collaboration. *Journal on Multimodal User Interfaces* 12: 77–89. <https://doi.org/10.1007/s12193-017-0250-2>.



## 2

# Communication Models for Remote Guidance

## 2.1 Introduction

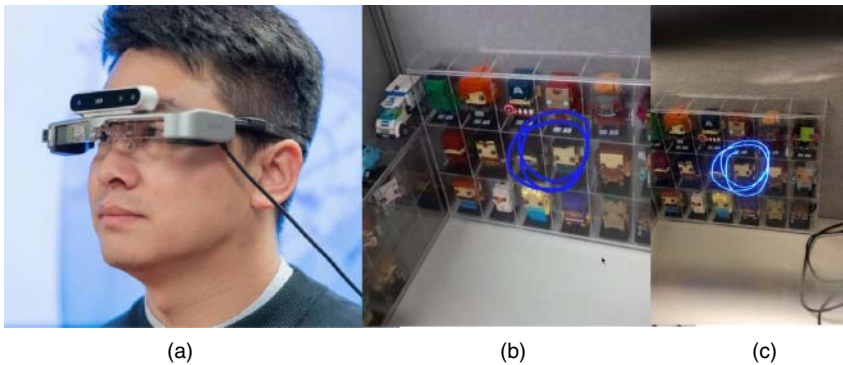
Communication is an essential part of remote collaboration, so understanding the impact of technology on communication is important in the system design process. For example, understanding how communication will change if one person cannot see what their remote collaborator is doing or if they have the ability to point or draw in their field of view. One way to do this is by using communication models; these are theoretical frameworks that can be used to predict communication behaviors when people use different collaboration technologies.

One important element of remote collaboration is to understand the impact of technology on communication behaviors. For example, Whittaker reviews using audio only to audio and video conferencing in a collaboration task and finds that people performed equally well but had very different communication patterns [1]. Previous researchers have developed a range of different communication models to explain how people communicate with one another and to predict the impact of technology on remote collaboration. For example, Clark and Brennan's grounding model of communication [2] has been used to predict communication behavior in video conferencing, especially when compared to audio-only conferencing [3].

In this book, our main focus is on Augmented Reality (AR), a collection of display, input, and tracking technologies that can be used to seamlessly overlay video imagery in the real world [4]. The ability to provide virtual visual and spatial cues makes AR an ideal technology for enhancing face-to-face and remote collaboration [5]. Previous collaborative AR systems have overlaid virtual video of remote collaborators in a user's real space [6], used shared virtual content to enhance face-to-face collaboration [7], and enabled a remote user to place virtual cues in a local person's workspace [8]. Studies with these systems have found that remote people feel a higher degree of social presence when using AR than when using video conferencing [6], they collaborate more naturally [9], and use behaviors similar to face-to-face collaboration [7].

In this chapter, we discuss how existing communication models can be used to predict the impact of different AR cues in remote collaboration and if a new communication model needs to be developed. Unfortunately, this is an area that has not been well-researched. Despite the potential of AR for remote collaboration, there are relatively few formal user studies conducted with collaborative AR systems. For example, In a survey of all AR user studies conducted between 2005 and 2014, Dey et al. [10] found less than 5% of studies involved collaborative systems, and very few of those collected communication measures. Marques et al. [11] suggested that there is “... minimal support of existing frameworks and a lack of theories and guidelines to guide the characterization of the collaborative process using AR.” So, there has been relatively little previous work done on exploring communication models in AR for remote collaboration, and there is a need for more research on this topic.

There are many different types of collaborative AR systems, but the focus of this chapter is specifically on head-worn AR systems for remote collaboration on physical tasks. A typical example is a system that uses a see-through head-mounted display (HMD) with a camera mounted on it that allows a local worker to stream a view of their workspace to a remote helper. The remote helper in turn can add virtual content to the local worker’s view to help assist them with the physical task that they are doing (see Figure 2.1). Figure 2.1a shows a typical version of such a system with a depth-sensing camera added to an Epson AR display. Figure 2.1b shows the view through the AR HMD and the remote expert view, where the expert is drawing on the live video feed to provide AR visual cues back into the local workers’ view. This type of system could be used in many applications, such as a remote expert helping a mechanic fix a car or an expert surgeon remotely assisting a novice doctor.



**Figure 2.1** A simple example of an HMD-based collaborative AR system. (a) The HMD with depth-sensing camera attached, (b) remote expert view with live annotation, (c) AR view.

There are many examples of research that have a similar setup, such as [6, 12–16]. This type of configuration is also becoming increasingly common in industrial applications. For example, Microsoft’s Remote Assist application uses the HoloLens2 AR HMD to allow a local worker to collaborate with remote helpers [17]. Remote Assist streams the HoloLens2 camera view to one or more remote users viewing the content on the web, who are then able to talk to the local worker, see what they are seeing, and place virtual arrows or other cues in the field of view.

Although not widely used, examples of systems like this are not new. Research on AR systems for remote collaboration dates back to the 1990s with the Shared-View work [18], and British Telecom’s CamNet system [19]. Since then, dozens of research papers have been published, but there have been relatively few studies of these systems from a communications perspective. Being able to evaluate this research from a communications perspective will help identify the research areas that should be further investigated, provide guidelines for improving the user experience, and establish the limitations of the current communication models. Just as using communication models improved video conferencing, the same type of approach could be used to improve AR systems for remote collaboration.

In this chapter, we review various communication models and show how they can be used to analyze communication in different AR interfaces for remote guidance on physical tasks. In the remainder of this chapter we first provide a historical review of communication models, especially focusing on remote communication (Section 2.2). Next, we show how communication models have been applied to analyze non-AR remote collaborative systems (Section 2.3) and research on the application of communication models to collaborative AR (Section 2.4). In Section 2.5, we discuss the limitations of current communication models and explore how they could be extended to accommodate all of the communication affordances of AR systems for remote assistance. Finally, we identify the research gaps that should be explored in the next generation of collaborative AR systems (Section 2.6).

The goal of this chapter is to provide the reader with enough understanding of communication models that they can use to predict the impact of various technology elements on AR systems for remote collaboration. This should enable them to develop better systems and to improve their own research in this area.

## 2.2 Overview of Communication Models

Communication theories attempt to describe and explain how people share knowledge and information with each other. Communication models are formalized concepts of the information-sharing process. They can be simple or complex and

there have been a wide variety of models developed. In this section, we provide a quick overview of some of the most important historical communication models.

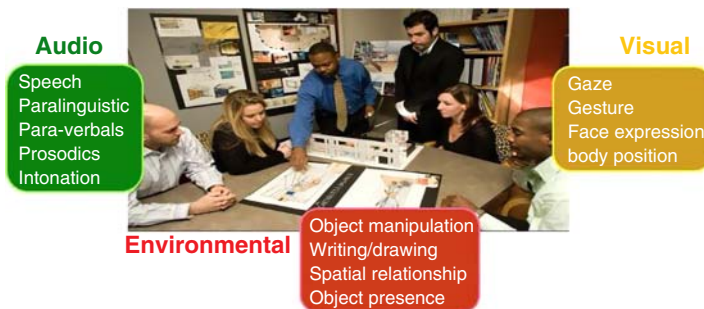
Formal models of communication date back thousands of years to Aristotle and his work on rhetoric [20]. In this classic work, he proposed a simple communication model with three parts: a speaker, a message, and a listener. Each of these parts is essential. For example, a speaker and their message do not communicate if there is no listener. These three elements of speaker, message, and listener have been used in many subsequent models, with Kumar noting that “Western theories and models of communication have their origin in Aristotle’s Rhetoric” ([21], p. 16).

In a similar way, Green et al. [22] present a simple human-to-human communication model that has three key components:

- The communication channels available.
- The communication cues provided by each of these channels.
- The affordances of the technology that affect the transmission of these cues.

They say that there are three main types of communication channels available: audio, visual, and environmental, where visual and audio cues are those that can be seen and heard, and environmental channels support interactions with the surrounding world. Depending on the communication medium, different communication cues may be able to be transmitted between collaborators. For example, using text chat will enable text messages to be sent between people but will prevent the communication of audio or environmental cues.

In face-to-face communication, a wide variety of communication cues are used when people collaborate together. These can be classified into Visual, Audio, and Environmental cues (see Figure 2.2). Audio cues include speech, paralinguistic, para-verbals, prosodics, intonation, and other types of audio. Visual cues are those generated by the user and include gaze, gesture, facial expression, and body position, among others. Finally, environmental cues include actions of the user in the environment to support communication, such as object manipulation, writing or



**Figure 2.2** Face-to-face communication cues.



**Figure 2.3** Introducing a separation between task space and communication space. (a) Face-to-face collaboration with task space a subset of communication space. (b) Remote collaboration with task space separate from the communication space.

drawing, object presence, and the spatial relationships between objects, among others. One of the goals of a communication model is to understand how variation in these elements can affect communication.

In addition to using a range of different communication cues, when people are collaborating on a task, there is a distinction between the task space and communication space (Figure 2.3a). The task space is the physical workspace that people are focusing on to complete a particular task, while the communication space is the space where people are able to see each other and share communication cues.

When people are collaborating face to face they can easily see each other and the range of different communication cues being used, so the task space is a subset of the communication space (Figure 2.3a). Ishii describes this as seamless communication because there is no functional separation between the task and communication space [23]. However, when people are working remotely, the task space is often separated from the communication space (see Figure 2.3b). For example, when video conferencing, people may have the face of their collaborator on one screen, while looking at a shared document on another. In this case, it is impossible to see many of the remote collaborator's communication cues at the same time as looking at the task space. So, there is an artificial seam between the communication space and task space. This is the type of impact of remote communication technology that needs to be predicted through communication models.

Green et al. [22] point out that the benefit of communication models is that they can be used to predict collaborative behavior and the impact of technology on collaboration. For example, if two people are talking over the phone, they are likely to use more verbal cues than if they were using a video conferencing link capable of sharing audio and visual cues. In the case of text-only communication, communication is reduced to one content-heavy visual channel, with a number of possible effects, such as less verbose communication, use of longer phrases, increased time to reach grounding, slower communication, and fewer interruptions.

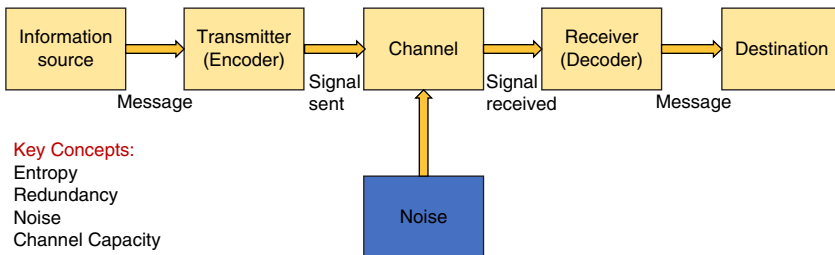
### 2.2.1 Linear Communication Models

Most communication models trace their roots back to Shannon and Weaver’s 1949 linear communication model [24]. Their model has a source, a transmitter, a signal, a receiver, and a destination (see Figure 2.4). Following Aristotle, the source is the equivalent of the speaker, and the destination is the same as the listener. Aristotle’s message gets converted into a signal at the transmitter. This signal is called a sent signal, and while it is being transmitted, noise is added to it, resulting in the received signal that reaches the receiver. For example, applying this model to a telephone call, the speaker’s voice is converted to an electrical signal conveyed over telephone lines, but the signal is degraded by additional noise in the telephone line that can make it difficult for the listener to hear.

The Shannon and Weaver communication model is unique as it was initially developed to describe communication over technology, namely telephones and radios. Shannon was focusing on the noise caused by the technology and correctly decoding the sender’s message. Although this model was designed for telecommunication, it has been widely used in other areas.

Around the same time, Berlo [25] developed a model that he described as “a model of the ingredients in communication.” It had four main parts: a source, a message, a channel, and a receiver (see Figure 2.5). The source and receiver were identical, with both having the same five characteristics: communication skills, attitudes, knowledge, social system, and culture. The message was composed of five elements: structure, content, treatment, and code, while the channel had the five senses: seeing, hearing, touching, smelling, and tasting.

Berlo believed that for effective communication to take place, the source and receiver had to be at the same level, such as having the same communication skills or similar knowledge. However, there are some limitations to this model, including not considering noise, having a lack of feedback, and it is a linear model. Most significantly, it assumes that people need to have the same knowledge or skill level for effective communication, which very rarely happens in everyday life.



**Figure 2.4** The Shannon and Weaver communication model.