Human–Computer Interaction Series

Human–Computer Interaction is a multidisciplinary field focused on human aspects of the development of computer technology. As computer-based technology becomes increasingly pervasive – not just in developed countries, but worldwide – the need to take a human-centered approach in the design and development of this technology becomes ever more important. For roughly 30 years now, researchers and practitioners in computational and behavioral sciences have worked to identify theory and practice that influences the direction of these technologies, and this diverse work makes up the field of human–computer interaction. Broadly speaking, it includes the study of what technology might be able to do for people and how people might interact with the technology.

In this series, we present work which advances the science and technology of developing systems which are both effective and satisfying for people in a wide variety of contexts. The human–computer interaction series will focus on theoretical perspectives (such as formal approaches drawn from a variety of behavioral sciences), practical approaches (such as the techniques for effectively integrating user needs in system development), and social issues (such as the determinants of utility, usability and acceptability).

Author guidelines: www.springer.com/authors/book+authors > Author Guidelines

For further volumes:
http://www.springer.com/series/6033

Alexander Waibel · Rainer Stiefelhagen
Editors

Computers in the Human Interaction Loop

Springer

Editors
Alexander Waibel
Universität Karlsruhe (TH)
Germany

Rainer Stiefelhagen
Universität Karlsruhe (TH)
Germany

ISSN 1571-5035
ISBN 978-1-84882-053-1 e-ISBN 978-1-84882-054-8
DOI 10.1007/978-1-84882-054-8
Springer Dordrecht Heidelberg London New York

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009921806

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Considerable human attention is expended on operating and attending to computers, and humans are forced to spend precious time fighting technological artifacts rather than doing what they enjoy and do well: human-human interaction and communication.

Instead of forcing people to pay attention to the artifact, that artifact should pay attention to humans and their interaction. Rather than forcing humans to work within a loop of computers, we would like to see computers serving in a Human Interaction Loop (CHIL). A CHIL computing environment aims to radically change the way we use computers. Rather than expecting a human to attend to technology, CHIL attempts to develop computer assistants that attend to human activities, interactions, and intentions. Instead of reacting only to explicit user requests, such assistants proactively provide services by observing the implicit human request or need, much like a personal butler would.

In 2004, a consortium of 15 laboratories from nine countries constituted itself as the CHIL Consortium to study this CHIL computing paradigm. Project CHIL, "Computers in the Human Interaction Loop," was supported as an Integrated Project (IP 506909) of the European Union under its 6th Framework Program and represents one of Europe's largest concerted efforts in the area of advanced human interfaces. It began on January 1, 2004, and ended successfully in August 2007. It was coordinated administratively by the Fraunhofer Institut für Informations- und Datenverarbeitung (IITB) and scientifically by the Interactive Systems Labs (ISL) at the University of Karlsruhe. The CHIL team includes leading research laboratories in Europe and the United States who collaborate to bring friendlier and more helpful computing services to society.

Instead of requiring user attention to operate machines, CHIL services attempt to understand human activities and interactions to provide helpful services implicitly and unobtrusively. To achieve this goal, machines must understand the human context and activities better; they must adapt to and learn from the human's interests, activities, goals, and aspirations. This requires machines to better perceive and understand all the human communication signals including speech, facial expressions, attention, emotion, gestures, and many more. Based on the perception and understanding of human activities and social context, innovative context-aware and proactive computer services became possible and were explored. Several prototype services were assembled and tested for performance and effectiveness: (1) the Connector (a proactive phone/communication device), (2) the Memory Jog (for supportive information and reminders in meetings), (3) collaborative supportive workspaces and meeting monitoring.

To realize the CHIL vision, the key research activities concentrate on four central areas:

- Perceptual technologies: Proactive, implicit services require a good description of human interaction and activities. This implies a robust description of the perceptual context as it applies to human interaction: Who is doing What, to Whom, Where and How, and Why. Unfortunately, such technologies – in all of their required robustness – do not yet exist. The consortium identified a core set of needed technologies and set out to build and improve them for use in CHIL services.
- Data collection and evaluation: Due to the inherent challenges (open environments, free movement of people, open distant sensors, noise, etc.), technologies are advanced under an aggressive R&D regimen in a worldwide evaluation campaign. In support of these campaigns, meeting, lectures and seminar data have been collected at more than five different sites, and metrics for the technology evaluations have been defined.
- Software infrastructure: A common and defined software infra-architecture serves to improve interoperability among the partners and offers a market-driven exchange of modules for faster integration.
- Services: based on the emerging technologies developed at different labs, using a common architecture, and within a user-centered design framework, CHIL services are assembled and evaluated. In this fashion, first prototypes are continually being (re-)configured and the results of user studies effectively exploited.

This book provides an in-depth discussion of the main research work and main results from an almost four-year effort on CHIL technologies and CHIL services. The work and insights reported represent a major advance in the state-of-the-art of modern context-aware computing and interface technologies.

A description of the CHIL framework, however, would not be complete without mentioning the excellent and constructive cooperation within the CHIL Consortium. The spirits of competitive cooperation as well as cooperative competition turned out to be key factors for the scientific achievements of CHIL.

July 2008

Alex Waibel, Rainer Stiefelhagen,
and the CHIL Consortium

The CHIL Consortium

Financial Coordinator	Prof. Hartwig Steusloff and Dr. Kym Watson Fraunhofer Institute IITB, Fraunhoferstr. 1 76131 Karlsruhe, Germany Email: steusloff@iitb.fraunhofer.de
Scientific Coordinator	Prof. Alex Waibel and Dr. Rainer Stiefelhagen Universität Karlsruhe (TH), Fakultät Informatik Am Fasanengarten 5, 76131 Karlsruhe, Germany Email: waibel@ira.uka.de, stiefel@ira.uka.de

Table 1. Contact persons.

Participant name	**Short name**	**Country**
Fraunhofer-Gesellschaft zur Förderung der Angewandten Forschung e.V. through its Fraunhofer Institut für Informations- und Datenverabeitung, IITB	FhG/IITB	Germany
Universität Karlsruhe (TH) through its Interactive Systems Laboratories (ISL)	UKA-ISL	Germany
DaimlerChrysler AG	DC	Germany
Evaluations and Language Resources Distribution Agency	ELDA	France
IBM Ceska Republika	IBM CR	Czech Republic
Research and Education Society in Information Technologies	RESIT (AIT)	Greece
Institut National de Recherche en Informatique et en Automatique through GRAVIR-UMR5527	INRIA[1]	France
Foundation Bruno Kessler - irst	FBK-irst	Italy
Kungl Tekniska Högskolan	KTH	Sweden
Centre National de la Recherche Scientifique through its Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI)	CNRS-LIMSI	France
Technische Universiteit Eindhoven	TUE	Netherlands
Universität Karlsruhe (TH) through its Institute IPD	UKA-IPD	Germany
Universitat Politècnica de Catalunya	UPC	Spain
The Board of Trustees of the Leland Stanford Junior University	Stanford	USA
Carnegie Mellon University	CMU	USA

Table 2. Partners in the CHIL Consortium.

[1] Institut National de Recherche en Informatique et en Automatique (INRIA) participated through GRAVIR-UMR5527, a joint laboratory of INRIA, Centre National de la Recherche

Acknowledgments

We gratefully acknowledge the financial support of the European Commission under the 6th Framework Program. Without the support of the EC and the efforts of the two heads of Unit E1, Nino Varile and Bernhard Smith, work in this area and this Integrated Project would not have been possible.

We would like to express our sincerest gratitude to the project officers at the EC, Mats Ljungqvist and Anne Bajart. As responsible representatives of the European Commission, they have accompanied the work of this IP at all times with great enthusiasm, encouragement, and competent advice and provided invaluable guidance and feedback, scientifically as well as administratively. From their guidance, a highly focused and disciplined, yet highly creative and innovative program resulted.

We would also like to thank the expert review panel for its valuable feedback and advice throughout the project: Josef Kittler, University of Cambridge (UK), Jean Vanderdonckt, Université Catholique de Louvain (Belgium), Michael Wooldridge, University of Liverpool (UK), and Hans Gellersen, Lancaster University (UK).

We also thank the members of the international advisory board of the CHIL project who provided numerous valuable comments and feedback early on, which streamlined the process and focused the Consortium's efforts. The board included Sadaoki Furui, Tokyo Institute of Technology (Japan), Martin Herman, NIST (USA), Joseph Mariani, Ministère délégué à la Recherche (France) at the time of CHIL, Matthew Turk, University of California, Santa Barbara (USA), Ipke Wachsmuth, University Bielefeld (Germany), and Seiichi Yamamoto, Advanced Telecommunications Research Institute International, Kyoto (Japan).

We would like to thank NIST, the National Insititute of Standards and Technology (USA), for its excellent cooperation. NIST included and supported CHIL's participation in its annual Rich Transcription (RT) evaluations during the entire duration of the project. Together with the NIST team, a new evaluation platform for the evaluation of multimodal perception technologies, CLEAR – Classification of Events, Activities and Relationships – was founded in 2006 and jointly organized in 2006 and 2007. We thank the NIST speech group for its cooperation, advice, and friendly support, especially John Garofolo, Rachel Bowers, Jonathan Fiscus, Audrey Le, Travis Rose, Martial Michel, Jerome Ajot, and Kathy Gallo. We also thank Vincent Stanford for his support in using NIST's Marc III microphone array in our smart rooms.

Last, but no means least, all credit is due to the efforts of the CHIL team. A select group from the best laboratories in Europe and the United States, the CHIL Consortium has brought together a remarkable group of research talent that in many ways achieved the enormous feat of demonstrating successful CHIL computing in a very short time. Driven by a joint vision, the CHIL researchers, students, and staff members worked way beyond the call of duty to make the program a success.

Only some of these appear as chapter authors of this book. However, there were more people, including administrative staff members, researchers, and students, who

Scientifique (CNRS), Institut National Polytechnique de Grenoble (INPG), and Université Joseph Fourier (UJF) provided invaluable contributions to this effort. We would thus like to acknowledge their help here:

The Interactive Systems Labs (ISL) of Universität Karlsruhe (TH) would gratefully like to acknowledge the help of the following persons: Annette Römer and Silke Dannenmaier did a terrific job in managing the bookkeeping of our lab in this large research project, making countless travel arrangements, and organizing various meetings, workshops and events. Margit Rödder did a wonderful job in organizing the dissemination activities of the project, including the organization of two technology days, the project's presentation at the IST fair in Helsinki (for which we won a best exhibition prize!), and numerous meetings and project workshops. She also handled the production of the project's "technology catalog," various presentations and publications, and all interactions with the press. The ISL computer vision team was greatly supported by a number of student members. These include Johannes Stallkamp (face recognition from video), Christian Wojek (audiovisual activity classification and tracking), Julius Ziegler (articulated body tracking), Hua Gao (3D face recognition), Mika Fischer (face recognition), Lorant Toth (open set face recognition), Florian Van De Camp (tracking with active cameras) and Alexander Elbs (multiperson tracking). Gopi Flaherty provided great help in programming smart telephones. Tobias Kluge helped evaluate the Connector service. We also would like to thank the following members of the speech recognition team: Florian Kraft, Sebastian Stüker, Christian Fügen, Shajit Ikhbal and Florian Metze. Furthermore, Kenichi Kumatani contributed to work on acoustic source separation and audiovisual speech recognition. The following students contributed to speech recognition and acoustic source localization: Tobias Gehrig, Uwe Maier, Emilian Stoimenov, Ulrich Klee and Friedrich Faubel (feature enhancement). We also want to thank Cedrick Rochet for his help with data collection, quality control, and distribution of data for the CLEAR workshops. We would also like to thank Kornel Laskowski for his contributions to acoustic emotion recognition and Dan Valsan for his technical support. Finally, we would like to thank Andrea Krönert and Anja Knauer for their help in incorporating the valuable suggestions of the copy editor into the final manuscript of this book.

Fraunhofer IITB and Universität Karlsruhe – ISL would not have been able to coordinate the large integrated R&D CHIL project without a professional consortium management team. We especially thank Tanja Schöpke (Fraunhofer contracts department), Henning Metzlaff (Fraunhofer international project finances), and Marika Hofmann (project management assistant) for their excellent contributions to the project. Special thanks are also due to Fernando Chaves, Falk Walter, and Sibylle Wirth for their creative ideas in establishing the CHIL project server and supporting the dissemination of information to the public. Fraunhofer IITB gratefully acknowledges the significant R&D work on the CHIL software architecture achieved through the contributions of (in alphabetical order) Stefan Barfuß, Thomas Batz, Laurenz Berger, Tino Bischoff, Gottfried Bonn, Ingo Braun, Ulrich Bügel, Axel Bürkle, Fernando Chaves, Daniel Dahlmeier, Niels Dennert, Rusi Filipov, Walter Gäb, Torsten Geiger, Torsten Großkopf, Rainer Helfinger, Jörg Henß, Alice Hertel, Reinhard Herzog, Fritz Kaiser, Ruth Karl, Carmen Kölbl, Thomas Kresken, Pascal Lang, Jürgen Moßgraber, Wilmuth Müller (leader of the IITB R&D work), Michael Okon, Dirk

Pallmer, Uwe Pfirrmann, Manfred Schenk, Gerd Schneider, Michael Stühler, Gerhard Sutschet (leader of the CHIL workpackage on architecture), Daniel Szentes, Martin Thomas, Christoph Vedder, Katrin Vogelbacher, Gero Volke, Patrick Waldschmitt, Heiko Wanning, Martin Wieser, Boxun Xi, Mustafa Yilmaz, and Tianle Zhang.

IBM would like to thank all CHIL partners for their enthusiasm and excellent collaboration during the entire project. Special thanks go to the architecture team (WP2) for its cooperation and many valuable comments during the development of the middleware tools SitCom and CHiLiX, as well as the perception technologies teams under the WP4 and WP5 workpackages of the project. IBM would also like to acknowledge students of the Czech Technical University (CTU) involved in student projects within IBM-CTU cooperation: Petr Slivoň, Tomáš Solár, and Karel Pochop (3D visualization in SitCom), Michal Pavelka (visualization of history in SitCom), and Robert Kessl (support for CHiLiX and SitCom). In addition, IBM will like to acknowledge the work of colleagues Vit Libal and Stephen Chu in data collection and speech technology development efforts, as well as Andrew Senior (formerly with IBM), Zhenqiu Zhang (formerly with the University of Illinois, Urbana-Champaign), and Ambrish Tyagi (formerly with the Ohio State University) for work on visual tracking technology. In particular, Ambrish Tyagi has been the main contributor to the IBM data collection efforts. Furthermore, Patrick Lucey (Queensland University of Technology) contributed to multipose AVASR work on CHIL data. Last, but not least, we would like to thank our management team, in particular David Nahamoo, Chalapathy Neti, Michael Picheny, and Roberto Sicconi for their continued support during various stages of the project.

The AIT CHIL team would like to acknowledge Fotios Talanztis, who has played a leading role in AIT's audio research, as well as Christos Boukis for his valuable contributions in the same field. Special thanks to our students Siamak Azodolmolky, Ghassan Karame, Panos Kasianidis, Nikos Katsarakis, Vassilis Mylonakis, Ippokratis Pandis, Elias Retzeperis, George Souretis, Kostas Stamatis, and Andreas Stergiou, who have undertaken the (real-time) implementation of AIT's audio and visual tracking technologies, as well as of several middleware components of the CHIL architecture and applications.

INRIA would like to gratefully thank the following people for their participation and substantial contributions to the CHIL project: Patrick Reignier (situation modeling, situation implementation), Dominique Vaufreydaz (audio processing, speech detection and analysis), Jerome Maisonnasse (human interaction analysis, user studies and experiments), Sebastien Pesnel (video tracking system, perceptual components and system architecture), Alba Ferrer-Biosca (3D video tracking system, perceptual components), Alban Caporossi (video tracking system), and Augustin Lux (image processing, perceptual components).

FBK-irst would like to thank the following people for their substantial contributions to the project: Koray Balci and Nadia Mana (embodied agents), Alberto Battocchi, Ilenia Graziola, Vera Falcon (user studies for the Social Supportive Workspace and the Relational Report), Luca Brayda and Claudio Bertotti (to Mark III bugs, design of the new array), Luca Cristoforetti (audio acquisition system, data acquisition),

Massimiliano Ruocco (model-based head orientation from optical flow), Piergiorgio Svaizer (advice on signal processing), Francesco Tobia (low-level vision software architecture, data acquisition, participation in the CLEAR evaluation campaigns), and Stefano Ziller and Sebastian Rizzo (vision-based motion capture).

KTH would like to thank the CHIL Consortium for an excellent joint effort and a well-organized and smoothly run project. The multimodal team was considerably strengthened by Magnus Nordstrand, Gunilla Svanfeldt, and Mikael Nordenberg. Inger Karlsson and Kåre Sjölander dedicated themselves to the emotion recognition task. Caroline Bergling provided excellent and timely administrative support during the whole project.

Many researchers in the Spoken Language Processing and some in Language, Information and Representation Group at LIMSI contributed to the CHIL project The main activities at LIMSI concerned the activities of "Who and Where" and "What", to which the following persons directly contributed: Gilles Adda, Martine Adda-Decker, Claude Barras, Eric Bilinski, Olivier Galibert, Jean-Jacques Gangolf, Jean-Luc Gauvain, Lori Lamel, Cheung-Chi Leung, Sophie Rosset, Holger Schwenk, Delphine Tribout, and Xuan Zhu. Early in the project, some initial work was carried out for emotion recognition, involving the contributions of Laurence Devillers, Laurence Vidrascu, Martine L'Esprit, and Jang Martel. We also would like to thank Alexandre Allauzen, Hélène Bonneau-Maynard, Michèle Jardino, Patrick Paroubek, and Ioana Vasilescu for their support and fruitful discussions, as well as Bianca Dimulescu-Vieru and Cecile Woehrling for their help when called upon. We also would like to thank the people in the background who contributed to the smooth functioning of the CHIL project: the administrative and technical staff at the laboratory and the CNRS regional headquarters. In particular, we would like to mention Martine Charrue, Béatrice Vérin-Chalumeau, Karine Bassoulet-Thomazeau, Joëlle Ragideau, Nadine Pain, Bernard Merienne, Elisabeth Piotelat, Annick Choisier, Pascal Desroches, and Daniel Lerin for their help throughout the project. We also acknowledge the support from the laboratory director, Patrick Le Quéré, and co-director Philippe Tarroux.

The Department of Industrial Design of the Technische Universiteit Eindhoven wishes to thank the following people: Adam Drazin, Jimmy Wang, Olga Kulyk, Rahat Iqbal, Anke Eyck, Kelvin Geerlings, Ellen Dreezens, and Olga van Herwijnen, all of whom were involved in the research for some time, Martin Boschman, Leon Kandeleers, and Eugen Schindler for technical support, and Richard Foolen and Nora van den Berg for administrative support. Furthermore, we would like to thank Herman Meines of the Governmental Department of VROM in The Hague for allowing us to make recordings of strategic meetings, and Alan Kentrow, Ernst Teunissen, Shani Harmon, and Neil Pearse of the Monitor Group for spending their time in numerous face-to-face and telephone meetings to discuss data analysis and to explore opportunities for applying CHIL technologies to the coaching of teamwork skills.

The Institute for Program Structures and Data Organization of Universität Karlsruhe (TH) appreciates the contributions of Yue Ying Zhi to the ontology visualization front end of the CHIL Knowledge Base Server. Thanks are also due to Stefan Mirevski and Ivan Popov for laying out the foundations of the formal specification of the CHIL OWL API. We wish to express our gratitude to Tuan Kiet Bui, who

devised the database back end of the CHIL Knowledge Base Server, implemented code generators for the CHIL OWL API, and contributed significantly to our site visit demonstrations. We are grateful to Guido Malpohl for his valuable advice and support in resolving networking issues during the development of the ChilFlow communication middleware. Special thanks go to the AIT CHIL team for their efforts in early adopting the CHIL Ontology.

TALP Research Center and the Image Processing Group of Universitat Politècnica de Catalunya (UPC) would like to acknowledge all groups and researchers who have contributed to the project activities. The authors are grateful to the CHIL partners who provided the evaluation seminar databases and also to all participants in the evaluations for their interest and contributions. We would like to thank the following individuals for their contributions to the success of the project: Joachim Neumann and Shadi El-Hajj for their contributions to the demos and the CHIL UPC movie; Albert Gil for the multicamera recording system in the UPC's smart room; Javier Ruiz and Daniel Almendro for the module wrappers and the solutions to real-time intermodule communication; Joan-Isaac Biel and Carlos Segura for the acoustic event detection system and the demo GUI. Carlos Segura is also thanked for helping to develop the isolated acoustic event database. We would like to acknowledge Jan Anguita and Jordi Luque for their work on speaker identification task, Pascual Ejarque and Ramon Morros for their contributions to multimodal person identification, Alberto Abad and Carlos Segura for their efforts on speaker tracking and orientation tasks, José Luis Landabaso and Jordi Salvador for their contributions to person tracking, Christian Ferran and Xavi Giró for their efforts in object detection, and Mihai Surdeanu for his contribution to question answering.

Stanford University would like to acknowledge the contributions from the following people: Anna Ho, Shailendra Rao, Robby Ratan, Erica Robles, Abhay Sukumaran, Leila Takayama, and Qianying Wang.

The Interactive Systems Laboratories at CMU would like to greatfully mention several lab members who helped and supported our work during the CHIL project: Jie Yang led efforts on video analysis and multimodal interaction; Datong Cheng assisted in collecting and analyzing video data; Tanja Schultz advised our developments in automatic recognition of meeting speech data at CMU as well as at Universität Karlsruhe (TH); Qin Jin worked on detecting of persons' identities from acoustic data; our data collection team supported the CHIL project in building a meeting room scenario, collecting meeting data, transcribing data, and labeling emotionally relevant behavior and instances of laughter. We would like to thank Nolan Pflug, Spike Katora, Joerg Brunstein (technical support and data collection), Timothy Notari, Julie Whitehead, Zachary Sloane, Matthew Bell, and Joseph P. Fridy (transcription and labeling) for their diligent work and great support.

Contents

Part III Services

List of Figures

List of Tables