CHIL: COMPUTERS IN THE HUMAN INTERACTION LOOP

Alex Waibe1[1], Hartwig Steusloff[2], Rainer Stiefelhagen[1] and the CHIL Project Consortium

[1]Interactive Systems Labs (ISL), University of Karlsruhe
[2]Fraunhofer Institut für Informations- und Datenverarbeitung (IITB)

ABSTRACT

CHIL ("Computers in the Human Interaction Loop") is an Integrated Project under the European Commission's Sixth Framework Programme. The CHIL consortium is jointly coordinated by Universität Karlsruhe (TH) and the Fraunhofer Institute IITB. CHIL was launched on January 1st, 2004.

The objective of this project is to explore and create environments in which computers serve humans who focus on interacting with other humans as opposed to having to attend to and being preoccupied by the machines themselves. Instead of computers operating in an isolated manner, and humans [thrust] in the loop [of computers]. CHIL puts Computers in the Human Interaction Loop (CHIL).

Fifteen partners from *nine* countries in Europe and the US collaborate in the CHIL Consortium to design Technologies and Computer Services that model humans and the state of their activities and intentions. A complete perceptual context enables a family of CHIL computing services that provide helpful assistance implicitly, requiring a minimum of human attention or interruptions.

1. PROJECT DESCRIPTION

The objective of the CHIL project is to create environments in which computers serve humans who focus on interacting with other humans as opposed to having to attend to and being preoccupied with the machines themselves. Instead of computers operating in an isolated manner, and Humans [thrust] in the loop [of computers], we will put Computers in the Human Interaction Loop (CHIL). We design Computer Services that model humans and the state of their activities and intentions. Based on the understanding of the human perceptual context, CHIL computers are enabled to provide helpful assistance implicitly, requiring a minimum of human attention or interruptions (see also the CHIL – Scenarios section).

To achieve this overall vision, a broad set of key scientific issues is proposed:

- Multimodal Perceptual User Interfaces that observe, recognize, fuse, and interpret all available cues and clues to explain human-human activities and intentions. Fundamental new algorithms are needed to achieve these capabilities (see the CHIL – Technologies section).
- A suite of Services that instantiate CHIL Computing based on perceptual context awareness and understanding of human activity. These services must balance implicit and explicit computer interaction, and must deliver information in an appropriate manner. Services include better ways of connecting people (without phone-tag), supporting human memory, & providing meeting support (see CHIL – Services) and more.
- A supportive infrastructure that supports CHIL Services including Automomic Computing, self-healing and self-maintaining software, flexible architecture, and a networked infrastructure integrating numerous devices intermittently and dynamically. The resulting shift from Human-Computer Interaction only (requiring full human attention) to increased reliance on human-human interaction is expected to lead to human productivity gains and reduced computer frustration (see CHIL – Software Architecture).

2. CHIL – SCENARIOS

A CHIL scenario is a situation in which people interact face to face with people, exchange information, collaborate to jointly solve problems, learn, or socialize, by using whatever means (speech/language, gestures, body posture, data in electronic format, slides, etc.) they choose. CHIL monitors the environment and provides useful services. During the three years of the project, we will focus on two scenarios: the office and the lecture room. The selection is motivated by a number of properties that make the two particularly challenging. It must be kept in mind, however, that the CHIL vision is more general, and that these are only concrete instantiations of a broader class of scenarios.

3. CHIL – SERVICES

Within this project we are going to implement several services that instantiate the paradigm of CHIL services. These will include the following services:

3.1. Memory Jog (MJ)

MJ helps the attendees by providing information related to the development of the event (meeting/lecture) and to the participants. MJ provides context- and content-aware information pull and push, both personalized and public.

3.2. Attention Cockpit (AC)

AC monitors the attention and interest level of participants, supporting individuals who want more or less involvement in the discussion. It can also inform the Socially-Supportive Workspaces about the attentional state of the participants.

3.3. Connector

Context-aware connecting services ensure that two parties are connected with each other at the right place, time and by the best media, when it is most appropriate and desirable for both parties to be connected.

3.4. Socially-Supportive Workspaces (SSW)

SSW is an infrastructure for fostering cooperation among participants, whereby the system provides a multimodal interface for entering and manipulating contributions from different participants, e.g., enabling joint discussion of Minutes, or joint accomplishment of a common task, with people proposing their ideas, and making them available on the shared workspace, where it is discussed by the whole group. The SSW provides a facilitator functionality that is able to monitor group activities to keep it on track, such as suggesting moving on the next task, and can better support social relationships.

4. CHIL – TECHNOLOGIES

The envisioned CHIL services, as outlined above, are not possible without a detailed understanding of the human state, human activities and intentions. Each of the services must know who is doing what with whom, why, and where in a given space, in order to consider appropriate actions. To answer these questions, we must continuously track human activities, using all perception modalities available, and build static and dynamic models of the scene. We must learn user profiles, detect behavioral patterns and be capable of combining the perceived multimodal information to better analyze the scene and provide pertinent assistance. In pursuing its target, CHIL will develop innovations advancing the state of the art in a wide range of component technologies:

- Multi-sensor, multi-modal processing for robust person localization, tracking and identification under unconstrained conditions (acoustic noise, visual occlusion, non-frontality, illumination variation)
- Body expression at various scales (body movements, gestures and postures)
- Focus of attention tracking with multiple moving persons
- Textual transcript generation for large vocabulary conversational speech with always on audio capturing under far-field conditions
- Audio-visual far-field speaker localization, source separation and speech recognition
- Multi-modal emotion identification by facial expression analysis and emotional voice classification
- Real-time topic identification with imperfect transcripts and integration with other cues such as initiative and conversational style etc.
- Audio-Visual Scene analysis for activity detection and description
- Context-aware personalized presentation
- Targeted audio
- Strategies for privacy and control management
- Architecture and infrastructure allowing for real-time processing of multi-sensor data, active recruitment and back-off of sensors and modalities, privacy and control management

Reliable output from these technologies will enable the realization of the services presented in the previous section. The component technologies developed in CHIL can be organized into the following three groups:

4.1. Who and Where

CHIL services will need to know the answer to the basic questions "Who" and "Where": First, who is the person or the group of people involved in the event(s)? And, second, where is this-and-that happening in the given environment? A third related question is "When": when was the such-and-such said/suggested, or done?, but technologies needed for answering this last question (time-stamping) are less challenging.

4.2. What

The next group of component technologies will target the detection of "What" is happening at any time, understanding what is being said, what is a conversation

about, what were the decisions, conclusions and action points. Furthermore, in order to engage proactively (or reactively) in the human communication loop, CHIL needs presentation technologies for the delivery of contents.

4.3. Why and How

Finally, CHIL will provide some insight into "Why" and in which way ("How") something has happened. This includes reasoning about the activities going on in a scene and about the attitudes of people when they interact with each other.

4.2. Abstraction levels

As we progress from one question to the next, we assume that the lowest level of scene and feature analysis will be targeted in the "Who and Where" section, the first level of abstraction related to scene interpretation is then left to the "What" section. Higher-level scene understanding is finally addressed under "Why and How". Obviously, the output from lower level processing (i.e. Who, Where or What) will be used as input for analysis tasks in higher processing levels (i.e. What, How and Why).

5. CHIL – SOFTWARE ARCHITECTURE

Among the objectives of the CHIL project is to specify and implement an architecture capable of supporting the implementation scenarios described earlier in this section. A main driver for devising the architecture is the need for supporting synergy and interoperation of individual components. Instead of providing an ad hoc integration of components, the architecture will provide the means for achieving integration in a structured fashion. Moreover, it will enable management of multimodal user interactions. Thus, in terms of software infrastructure the architecture will support components communication and multimodal interactions. Based on this infrastructure, strategies for situation detection, assessment and decision-making will be implemented.

5. CONCLUSION

CHIL "Computers in the Human Interaction Loop" is an "Integrated Project" effort funded by the European Union under the 6th Framework Program.
The CHIL team is a consortium of internationally renowned research labs in Europe and the US, who collaborate to bring friendlier and more helpful computing services to society. Rather than requiring user attention to operate machines, CHIL services attempt to understand human activities and interactions to provide helpful services implicitly and unobtrusively.

CHIL represents a vision of the future - a new approach to more supportive and less burdensome computing and communication services, and state of the art multimodal and perceptual user interface technologies in European Union and the US aiming at this vision. The team sets out to study the technical, social and ethical questions that will enable this next generation of computing in a responsible manner.

For further information please contact:
CHIL Scientific Coordinator:
Universität Karlsruhe (TH)
Interactive Systems Labs
http://isl.ira.uka.de
Prof. A. Waibel, ahw@cs.cmu.edu
Dr. Rainer Stiefelhagen, stiefel@ira.uka.de

CHIL Financial Coordinator:
Fraunhofer Institute for Information and Data Processing
http://www.iitb.fhg.de
Prof. H. Steusloff, hst@iitb.fhg.de
Dr. Kym Watson, wat@iitb.fhg.de

The CHIL consortium consist of the following partners:

DaimlerChrysler AG, Group Dialogue Systems, http://www.daimlerchrysler.com, Dr. Klaus Linhard, Germany

- ELDA, Evaluations and Language resources Distribution Agency, http://www.elda.fr, Dr. Khalid Choukri, France
- IBM Ceska Republika, http://www.cz.ibm.com, Jan Sedivy, Czech Republic
- RESIT, Research and Education Society in Information Technologies, http://www.ait.gr, Prof. Lazaros Polymenakos, Greece
- INRIA (Institut National de Recherche en Informatique et en Automatique), Project GRAVIR, Prof. James Crowley, http://www.inrialpes.fr, France
- IRST (Instituto Trentino di Cultura), http://www.itc.it, Dr. Fabio Pianesi, Italy
- KTH (Kungl Tekniska Högskolan), http://www.speech.kth.se, Prof. Rolf Carlson, Sweden.
- CNRS, LIMSI (Centre National de la Recherche Scientifique through its Laboratoire d'Informatique pour la mécanique et les sciences de l'ingénieur), http://www.limsi.fr, Dr. Lori Lamel, France

- TUE (Technische Universiteit Eindhoven), http://www.tue.nl, Dr. Jacques Terken, The Netherlands
- IPD, Universität Karlsruhe (TH) through its Institute IPD, http://www.ipd.uka.de, Prof. Walter Tichy, Germany
- UPC, Universitat Politècnica de Catalunya, http://www.upc.es, Prof. Josep R. Casas, Spain
- Universität Karlsruhe (TH), Interactive Systems Labs, http://isl.ira.uka.de, Prof. Alexander Waibel, Karlsruhe, Germany
- Fraunhofer Institut für Informations- und Datenverarbeitung (IITB), http://www.iitb.fraunhofer.de, Prof. Hartwig Steusloff, Karlsruhe, Germany
- Stanford University, http://www.stanford.edu, Prof. Clifford Nass, USA
- CMU, Carnegie Mellon University, http://www.is.cs.cmu.edu, Prof. Alexander Waibel, USA

6. REFERENCES

[1] CHIL project web-site: http://chil.server.de/