CHIL - Computers in the Human Interaction Loop - Electronic Butlers simplify your daily life

Margit Rödder

A cell phone ringing in the theatre, trying for hours to reach someone on the phone, attending a meeting and forgetting the documents, endless discussions, forgetting the name of a partner or friend… - that will soon be history. CHIL provides useful proactive and intelligent services, which will cause a fundamental shift in the way we use computers today. We aim to realize computer services that are delivered to people in an implicit, indirect and unobtrusive way. Computers in the Human Interaction Loop (CHIL) aims to introduce computers into a loop of humans interacting with humans, rather than condemning a human to operate in a loop of computers. This will give humans the most valuable gift: more time.

CHIL Scenario

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, using all their natural ways of face to face communication (speech/language, gestures, body posture, etc.). Therefore the focus in the project is on two scenarios: offices and lecture rooms.

CHIL Services

In order to provide really useful proactive and intelligent services, the Who, Where, What, Why and How of Human activities and communication needs to be perceived and understood. This presents a fundamental departure from the way we have thought of user interfaces up till now. It requires robust, multimodal perceptual interfaces capable of tracking, identifying, recognizing and understanding the role, purpose and content of human communication, activities, state and their environment. If these machines in human contexts were available, a new class of digital services could be developed that would take concrete advantage of these new capabilities. CHIL explores four particular services as instantiations of this vision:

Memory Jog

The Memory Jog provides attendees with information related to a situation (e.g. a meeting or a lecture) and related to the participants in it. It provides context- and content-aware information pull and push, both personalized and public.

Relational Report

This report evaluates the individual's contribution to the group's activity. Multimedia reports about the relational behaviour of each participant are privately delivered as part of an automatic coaching system. The idea is that the whole organization might benefit from an increase of awareness of participants about their own behaviours during group activities.

Connector

The Connector is a context-aware connecting service ensuring that two parties get connected by the most appropriate media at the right time and place. Based on the observed context, and each party's preference, it decides when and how it is most appropriate and desirable for both parties to be connected.

Socially-Supportive Workspaces

Socially-Supportive Workspaces are 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 they are discussed by the whole group. The Socially-Supportive Workspaces provide a facilitator functionality that is able to monitor group activities to keep it on track, such as suggesting moving on to

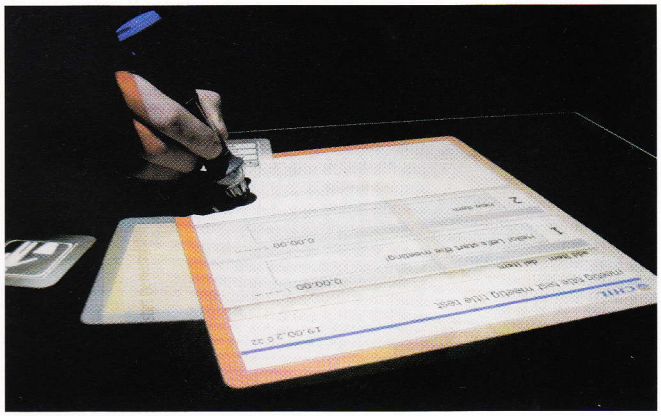

Figure 1: The Collaborative worksspace supports synchronous cooperation and participation by providing a shared information space and tools for managing face-to-face meeting.

Figure 2: The Connector Devices bring two parties together at the most appropriate time.

the next task, and can better support social relationships.

CHIL Technologies

In order to develop the described services, it is necessary to continuously track human activities, using all perception modalities available, and build static and dynamic models of the scene. With the different technologies, user profiles must be learned and behavioural patterns detected. The perceived multimodal information must be combined to better analyze the scene and to provide pertinent assistance.

In pursuing its target, the CHIL technique develops innovations advancing the state of the art in a wide range of component technologies:

- Audio Visual Person Tracking
- Pointing Gesture Recognition
- Face Detection and Face Recognition
- Head Pose Estimation
- Collaborative Workspace for Meetings
- Automatic Meeting Summarizer
- Agent Based Software Architecture
- Context Aware Management of Communication
- Animated Secretary - improved communication using virtual talking heads
- "SitCom" - the tool for situation modelling and visualization
- "AVASR" - the technology prototype for audio visual automatic speech recognition using far-field or close-talking audio-visual sensors
- "targeted audio", an array of small ultrasound speakers, that can deliver a very focussed audio beam.

CHIL - Software Architecture

Realizing the goal of the project demands that perceptual interfaces are integrated according to the design, purpose and objectives of the targeted services. Rather than focusing on an ad-hoc implementation of particular services, the CHIL project proceeds by specifying a structured method for interfacing with sensors, integrating technology components, processing sensorial input and ultimately composing non-obtrusive services as collections of basic service capabilities. Moreover, it enables management of multimodal user interactions. Thus, in terms of software infrastructure the architecture supports components communication and multimodal interactions. Based on this infrastructure, strategies for situation detection, assessment and decision-making are implemented.

Outlook

CHIL shows a vision of the future - a new approach to more supportive and less burdensome computing and communication services. The international and multidisciplinary team sets out to study the technical, social and ethical questions that will enable this next generation of computing in a responsible manner.

CHIL-PARTNERS

Fifteen partners from nine countries in Europe and the US collaborate in the CHIL consortium:

- Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB), Germany
- Universität Karlsruhe (TH), Interactive Systems Labs (ISL), Germany
- DaimlerChrysler AG, Group Dialogue Systems, Germany
- Evaluations and Language resources Distribution Agency (ELDA), France
- IBM Ceska Republika, Czech Republic
- Research and Education Society in Information Technologies (RESIT), Greece
- Institut National de Recherche en Informatique et en Automatique (INRIA), Lab GRAVIR, France
- Istituto Trentino di Cultura (IRST), Italy
- Kungl Tekniska Högskolan (KTH), Sweden

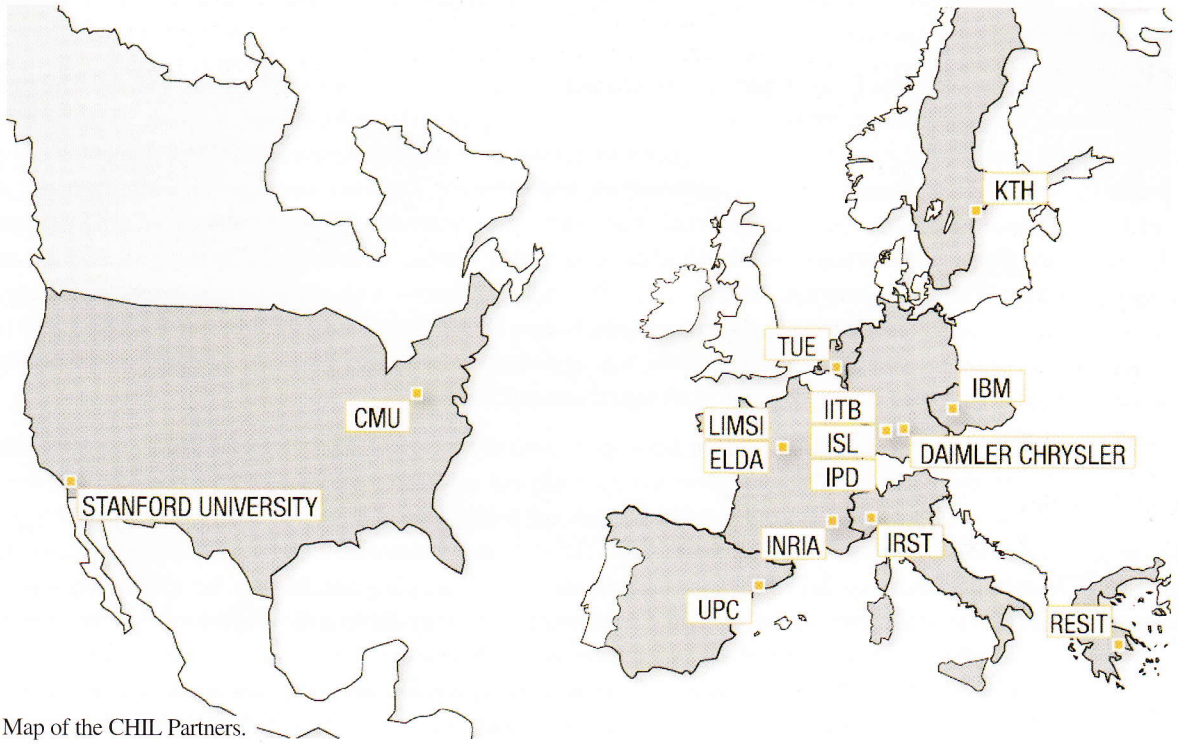

Figure 3: Map of the CHIL Partners.

- Centre National de la Recherche Scientifique, (CNRS), LIMSI, France
- Technische Universiteit Eindhoven (TUE), The Netherlands
- Universität Karlsruhe (TH), Institute for Program Structures and Data Organisation, (IPD), Germany
- Universitat Politècnica de Catalunya (UPC), Spain
- Stanford University, USA
- Carnegie Mellon University (CMU), USA

This Integrated Project CHIL IP 506909 is supported by funding in the thematic area Information Society Technologies under the Sixth Research Framework Programme of the European Union.

Further information under:

http://chil.server.de

Scientific Coordinator
Universität Karlsruhe (TH)
Interactive Systems Labs
http://isl.ira.uka.de
Prof. Alex Waibel, ahw@cs.cmu.edu
Dr. Rainer Stiefelhagen, stiefel@ira.uka.de

New Resources

ELRA-S0191 ZipTel

The ZipTel telephone speech database contains recordings of people applying for a SpeechDat prompt sheet via telephone. For the SpeechDat data collection, calls for participation were published in "phone", the customer magazine of the mobile telephone provider "e-plus", and in numerous newspapers all over Germany. In these calls, a telephone number was given where callers could order a SpeechDat prompt sheet. The calls were recorded by an automatic telephone server; callers were asked to provide address, ZIP code, city and telephone number.

Total number of recordings: 7746

Total duration: 14h

Format: SpeechDat Exchange Format, SAM, BAS Partitur Format (BPF)

	ELRA members	Non-members
For research use	627.17 Euro	754.35 Euro
For commercial use	4,627.17 Euro	4,754.35 Euro

GlobalPhone databases

GlobalPhone is a multilingual speech and text database collected at Karlsruhe University, Germany. The GlobalPhone corpus provides transcribed speech data for the development and evaluation of large vocabulary continuous speech recognition systems in the most widespread languages of the world. GlobalPhone is designed to be uniform across languages with respect to the amount of text and audio per language, the audio data quality (microphone, noise, channel), the collection scenario (task, setup, speaking style etc.), and the transcription conventions. As a consequence, GlobalPhone supplies an excellent basis for research in the areas of (1) multilingual speech recognition, (2) rapid deployment of speech processing systems to new languages, (3) language and speaker identification tasks, as well as (4) monolingual speech recognition in a large variety of languages.

To date, the GlobalPhone corpus covers 15 languages Arabic (Modern Standard Arabic), Chinese-Mandarin, Chinese-Shanghai, Croatian, Czech, French, German, Japanese, Korean, Portuguese (Brazilian), Russian, Spanish (Latin American), Swedish, Tamil, and Turkish. This selection covers a broad variety of language peculiarities relevant for Speech and Language Research and Development. It comprises widespread languages (Arabic, Chinese, Spanish), contains economically and politically important languages (Korean, Japanese, Arabic), and spans over wide geographical areas (Europe, America, Asia). The spoken speech covers a wide selection of phonetic characteristics, e.g. tonal sounds (Mandarin, Shanghai), pharyngal sounds (Arabic), consonantal clusters (German), nasals (French, Portuguese), palatized sounds (Russian), and more. The written language contains large orthographic variations, such as phonologic scripts (alphabetic scripts such as Roman, Cyrillic, Arabic; syllable-based scripts like Japanese Kana, Korean Hangul), and ideographic scripts (Chinese Hanzi and Japanese Kanji). The languages cover many morphological variations, e.g. agglutinative languages (Turkish, Korean), compounding languages (German), and also include scripts that completely lack word segmentation (Chinese).

The data acquisition was performed in countries where the language is officially spoken. In each language about 100 adult native speakers were asked to read 100 sentences. The read texts were selected from national newspaper articles available from the web to cover a wide domain with large vocabulary. The articles report national and international political news, as well as economic news mostly from the years 1995-1998. The speech data was recorded with a Sennheiser 440-6 close-speaking microphone and is available in identical characteristics for all languages: PCM encoding, mono quality, 16-bit quantization, and 16 kHz sampling rate. Most of the speech data was recorded in a quiet office, some are recorded in apartments, i.e. living room. The transcriptions are available in the original script of the corresponding language. In addition, all transcriptions have been romanized, i.e. transformed into Roman script applying customized mapping algorithms. The transcripts are validated and supplemented by special markers for spontaneous effects like stuttering, false starts, and non-verbal effects such as breathing, laughing, and hesitations. Speaker information, such as age, gender, occupation, etc. as well as information about the recording

setup complement the database. The entire GlobalPhone corpus contains over 300 hours of speech spoken by more than 1500 native adult speakers. The data are divided in speaker disjoint sets for training, development, and evaluation (80:10:10) and are organized by languages and speakers.

The list of available GlobalPhone resources is given below:

ELRA-S0192 GlobalPhone Arabic

The Arabic corpus was produced using the Assabah newspaper. It contains recordings of 78 speakers (35 males, 43 females) recorded in Tunisia, Palestine and Jordan. The following age distribution has been obtained: 20 speakers are below 19, 35 speakers are between 20 and 29, 13 speakers are between 30 and 39, 6 speakers are between 40 and 49, and 4 speakers are over 50.

ELRA-S0193 GlobalPhone Chinese-Mandarin

The Chinese-Mandarin corpus was produced using the Peoples Daily newspaper. It contains recordings of 132 speakers (64 males, 68 females) recorded in Beijing, Wuhan and Hekou, China. The following age distribution has been obtained: 16 speakers are below 19, 96 speakers are between 20 and 29, 16 speakers are between 30 and 39, 3 speakers are between 40 and 49 (1 speaker age is unknown).

ELRA-S0194 GlobalPhone Chinese-Shanghai

The Chinese-Shanghai corpus was produced using the Peoples Daily newspaper. It contains recordings of 41 speakers (16 males, 25 females) recorded in Shanghai, China. The following age distribution has been obtained: 1 speaker is below 19, 2 speakers are between 20 and 29, 13 speakers are between 30 and 39, 14 speakers are between 40 and 49, and 11 speakers are over 50.

ELRA-S0195 GlobalPhone Croatian

The Croatian corpus was produced using the HRT and Obzor Nacional newspapers. It contains recordings of 94 speakers (38 males, 56 females) recorded in Zagreb, Croatia, and parts of Bosnia. The following age distribution has been obtained: 21 speakers are below 19, 30 speakers are between 20 and 29, 14 speakers are between 30 and 39, 15 speakers are between 40 and 49, and 13 speakers are over 50 (1 speaker age is unknown).

ELRA-S0196 GlobalPhone Czech

The Czech corpus was produced using the Ceskomoravsky Profit Journal and Lidove Noviny newspaper. It contains recordings of 102 speakers (57 males, 45 females) recorded in Prague, Czech Republic. The following age distribution has been obtained: 16 speakers are below 19, 70 speakers are between 20 and 29, 2 speakers are between 30 and 39, 9 speakers are between 40 and 49, and 5 speakers are over 50.

ELRA-S01967 GlobalPhone French

The French corpus was produced using Le Monde newspaper. It contains recordings of 100 speakers (49 males, 51 females) recorded in Grenoble, France. The following age distribution has been obtained: 3 speakers are below 19, 52 speakers are between 20 and 29, 16 speakers are between 30 and 39, 13 speakers are between 40 and 49, and 14 speakers are over 50 (2 speakers age is unknown).

ELRA-S0198 GlobalPhone German

The German corpus was produced using the Frankfurter Allgemeine und Sueddeutsche Zeitung newspaper. It contains recordings of 77 speakers (70 males, 7 females) recorded in Karlsruhe, Germany. No age distribution is available.

ELRA-S0199 GlobalPhone Japanese

The Japanese corpus was produced using the Nikkei Shinbun newspaper. It contains recordings of 149 speakers (104 males, 44 females, 1 unspecified) recorded in Tokyo, Japan. The following age distribution has been obtained: 22 speakers are below 19, 90 speakers are between 20 and 29, 5 speakers are between 30 and 39, 2 speakers are between 40 and 49, and 28 speakers are over 50 (2 speakers age is unknown).

ELRA-S0200 GlobalPhone Korean

The Korean corpus was produced using the Hankyoreh Daily News. It contains recordings of 100 speakers (50 males, 50 females) recorded in Seoul, Korea. The following age distribution has been obtained: 7 speakers are below 19, 70 speakers are between 20 and 29, 19 speakers are between 30 and 39, and 3 speakers are between 40 and 49 (1 speaker age is unknown).

ELRA-S0201 GlobalPhone Portuguese (Brazilian)

The Portuguese (Brazilian) corpus was produced using the Folha de Sao Paulo newspaper. It contains recordings of 102 speakers (54 males, 48 females) recorded in Porto Velho and Sao Paulo, Brazil. The following age distribution has been obtained: 6 speakers are below 19, 58 speakers are between 20 and 29, 27 speakers are between 30 and 39, 5 speakers are between 40 and 49, and 5 speakers are over 50 (1 speaker age is unknown).

ELRA-S0202 GlobalPhone Russian

The Russian corpus was produced using the Ogonyok Gaseta and Express-Chronika newspapers. It contains recordings of 115 speakers (61 males, 54 females) recorded in Minsk, Belarus. The following age distribution has been obtained: 9 speakers are below 19, 76 speakers are between 20 and 29, 9 speakers are between 30 and 39, 15 speakers are between 40 and 49, and 6 speakers are over 50.

ELRA-S0203 GlobalPhone Spanish (Latin American)

The Spanish (Latin America) corpus was produced using the La Nacion newspaper. It contains recordings of 100 speakers (44 males, 56 females) recorded in Heredia and San Jose, Costa Rica. The following age distribution has been obtained: 20 speakers are below 19, 54 speakers are between 20 and 29, 13 speakers are between 30 and 39, 5 speakers are between 40 and 49, and 8 speakers are over 50.

ELRA-S0204 GlobalPhone Swedish

The Swedish corpus was produced using the Goeteborgs-Posten newspaper. It contains recordings of 98 speakers (50 males, 48 females) recorded in Stockholm and Vaernamo, Sweden. The following age distribution has been obtained: 9 speakers are below 19, 50 speakers are between 20 and 29, 12 speakers are between 30 and 39, 11 speakers are between 40 and 49, and 16 speakers are over 50.

ELRA-S0205 GlobalPhone Tamil

The Tamil corpus was produced using the Thinaboomi Tamil Daily newspaper. It contains recordings of 47 speakers (gender unspecified) recorded in India. No age distribution is available.

ELRA-S0206 GlobalPhone Turkish

The Turkish corpus was produced using the Zaman newspaper. It contains recordings of 100 speakers (28 males, 72 females) recorded in Istanbul, Turkey. The following age distribution has been obtained: 30 speakers are below 19, 30 speakers are between 20 and 29, 23 speakers are between 30 and 39, 14 speakers are between 40 and 49, and 3 speakers are over 50.

PRICES

• For S0194

	ELRA members	Non-members
For research use	300 Euro	355 Euro
For commercial use	1,800 Euro	2,125 Euro

• For S0192, S0193, S0195, S0196, S0197, S198, S0199, S0200, S0201, S202, S203, S204, S0206

	ELRA members	Non-members
For research use	600 Euro	700 Euro
For commercial use	3,000 Euro	3,600 Euro

• For S0205

	ELRA members	Non-members
For research use	100 Euro	125 Euro
For commercial use	500 Euro	600 Euro

Special prices for a purchase of several GlobalPhone Languages

• 5 Languages

	ELRA members	Non-members
For research use	2,600 Euro	3,000 Euro
For commercial use	13,500 Euro	16,200 Euro

• 10 Languages

	ELRA members	Non-members
For research use	5,000 Euro	6,000 Euro
For commercial use	24,000 Euro	28,800 Euro

• 15 Languages

	ELRA members	Non-members
For research use	7,500 Euro	9,000 Euro
For commercial use	31,500 Euro	37,800 Euro