# Speech Processing in Support of Human-Human Communication

Alex Waibel
*Carnegie Mellon University, USA and University of Karlsruhe, Germany*
*waibel@cs.cmu.edu*

Computers have become an essential part of modern life, providing services in a multiplicity of ways. Access to these services, however, comes at a price: human attention is bound and directed toward a technical artifact in a human-machine interaction setting at the expense of time and attention for other humans. This paper explores a new class of computer services that support human-*human* interaction and communication *implicitly* and *transparently*. Computers in the Human Interaction Loop (CHIL), require consideration of all communication modalities, multimodal integration and more robust performance. We review the technologies and several CHIL services providing human-human support. Among them, we specifically highlight advanced computer services for *cross-lingual* communication.

It is a common experience in our modern world, for humans to be overwhelmed by the complexities of technological artifacts around us, and by the attention they demand. While technology provides wonderful support and helpful assistance, it also gives rise to an increased preoccupation with technology itself and with a related fragmentation of attention. But as humans, we would rather attend to a meaningful dialog and interaction with other humans, than to control the operations of machines that serve us. The cause for such complexity and distraction, however, is a natural consequence of the flexibility and choices of functions and features that the technology has to offer. Thus flexibility of choice and the availability of desirable functions are in conflict with ease of use and our very ability to enjoy their benefits. The artifact cannot yet perform autonomously and requires precise specification of the machine's behavior. Standardization, better graphical user interfaces, multimodal human-machine dialog systems, speech, pointing, mousing have all contributed to improve the interface, but still force the user to interact with a machine at the detriment of other human-human interaction. To change the limitations of present day technology, machines must engage implicitly and indirectly in a world of humans, that is we must put Computers in the Human Interaction Loop (CHIL), rather than the other way round. Computers should assist humans engaged in human-human interaction, by providing implicit and proactive support. If technology could be "CHIL enabled" in this way, a host of new services could potentially be possible. Could two people be connected with each other at the best moment over the most convenient and best media, without phone tag, embarrassing ring tones and interruptions? Could an attendee in a meeting be reminded of participants' names and affiliations at the right moment without messing with a contact directory? Can meetings be supported, moderated and coached without technology getting in the way? And: Could computers enable speakers of different languages communicate and listen to each other gracefully across the language divide? Human assistants often provide such services; they work out logistical support, reminders, helpful assistance, and language mediation; they can do it reliably, transparently, tactfully, sensitively and diplomatically. Why not machines? Clearly, a lack of recognition and understanding of human activities, needs and desires are to blame, and an absence of socially adept computing services that mediate rather than intrude. In the following we focus on these two elements: 1.) technologies to track and understand the human context, and 2.) computing services, that mediate and support human-human interaction.

In contrast to classical human-machine interfaces, implicit computer support for human-human interaction requires a perceptual user interface with much greater performance, flexibility and robustness, than is available today. This challenge has lead to research aimed at tracking all the communication modalities in realistic recording conditions, rather than individual modalities in idealized recording conditions. CHIL and AMI, both Integrated projects under the 6th Framework Program of the European Commission, as well as CALO, a DARPA program are among the more recent efforts aiming to take on this challenge. In the following we will discuss computer services that support human-human interaction. To realize this goal, work concentrates on four key areas: The creation of robust perceptual technologies able to acquire rich and detailed knowledge about the human interaction context; the collection and annotation of realistic, audio-visual meeting and seminar data necessary for the development and systematic evaluation of such; the definition of a common software architecture to support reusability and exchangeability of services and technology modules; the implementation of a number of prototypical services offering proactive, implicit assistance based on the gained awareness about human interactions.

978-0-7695-3433-6/08 $25.00 © 2008 IEEE
DOI 10.1109/ISUC.2008.78