Guest Editorial
Special Issue on Human Computing

WE HAVE entered an era of enhanced digital connectivity. Computers and the Internet have become so embedded in the daily fabric of people's lives that people simply cannot live without them. We use this technology to work, to communicate, to shop, to seek out new information, and to entertain ourselves. In other words, computers are becoming full *social actors* that need to interact with people as seamlessly as possible. The key to development of computers as such social actors is to design human–computer interaction (HCI) that is *human centered*, built for humans based on human behavior models [1], [2]. In other words, HCI designs should focus on the human portion of the HCI context rather than on the computer portion, as was the case in classic HCI designs such as direct manipulation and delegation. They should transcend the traditional keyboard and mouse to include natural humanlike interactive functions, including understanding and emulating certain human behaviors such as affective and social signals. Hence, the key issue in the design of *human computing* (i.e., human-centered technologies capable of seamless interaction with people) is realizing machine understanding of human communicative behavior.

Machine understanding of human communicative behavior is inherently a multidisciplinary enterprise involving different research fields, including psychology, linguistics, computer vision, signal processing, and machine learning. There is no doubt that the progress in machine understanding of human interactive behavior is contingent on the progress in the research in each of those fields. As discussed by the Guest Editors of this Special Issue in [1] and [2], the main scientific and engineering challenges related to the realization of machine sensing and understanding of human communicative behavior include sensing of *what* is communicated (e.g., linguistic message, nonlinguistic conversational signal, and emotion), *how* the information is communicated (the person's facial expression, head movement, tone of voice, and hand and body gestures), *why*, i.e., in which context, the information is passed on (where the user is, what his current task is, and how he/she feels), and *which* (re)action should be taken to satisfy user needs and requirements. Human-centered technologies able to detect these subtleties of and changes in the user's communicative behavior and to initiate interactions based on this information, rather than simply responding to the user's commands, are the kind of technologies which could be called *human computing*.

Although the research on human computing is still in its pioneering phase, promising approaches have been recently reported on context sensing [2], machine analysis of human affective and social behavior [3]–[5], smart environments [6], and multimodal interfaces [7]. This Special Issue focuses on these topics, which form the essence of human computing. It provides a state-of-the-art overview of new paradigms and challenges in research on human computing and highlights the importance of this topic.

Most of the papers in this Special Issue focus on two challenging issues in human computing, namely, machine analysis of human behavior in group interactions and context-sensitive modeling. Past research on human interactive behavior has mostly focused on dyadic interactions, i.e., dialogues involving only two persons or one person and a virtual character. Intragroup interactions have been studied much less [5]. The main reason for this is the fact that group interactions are much more challenging, particularly from the technological point of view, as algorithms need to be developed capable of taking into account multiple signal sources involved in complex interaction patterns. The first five papers in the Special Issue deal with various aspects of intergroup interactions. Similarly, although context plays a crucial role in understanding human behavioral signals, since they are easily misinterpreted if the information about the situation in which the shown behavioral cues have been displayed is not taken into account, past efforts on human behavior understanding are usually context insensitive [2]. All papers in the Special Issue address some of context-related issues. Some papers propose methodology to answer one or more, i.e., *where*, *what*, *when*, *who*, *why*, and *how* (W5+), context-related questions. For example, the paper by Talantzis *et al.* [8] addresses the *who* question; the paper by Ba and Odobez [9] focuses on the *what* question, and the paper by Gunes and Piccardi [13] discusses a way of dealing with the *why* and *how* questions. Other papers report on methods for context modeling (e.g., papers by Dai *et al.* [10], Olguin Olguin *et al.* [11], and Brdiczka *et al.* [12]).

More specifically, Talantzis *et al.* [8] discuss the use of multiple acoustic and video sensors that allow audiovisual speaker tracking in cluttered indoor environments. In their experiments, they consider a meeting environment where there are discussions, coffee breaks, presentations, and participants leaving and entering the room. They employ an online acoustic source location system that uses time-delay estimations between microphones to estimate the actual speaker location. Added to this approach is the use of multiple synchronized cameras that try to distinguish the people that are present in the monitored

The work of Maja Pantic and Anton Nijholt is supported in part by the EU IST Programme Project FP6-0027787 (AMIDA) and in part by the EC FP7 Programme (FP7/2007-2013) under Grant 211486 (SEMAINE). The work of Maja Pantic is further supported by the European Research Council under the ERC Starting Grant ERC-2007-StG-203143 (MAHNOB).
Digital Object Identifier 10.1109/TSMCB.2008.2008372

space. There are no attempts to localize the speaker using characteristics of nonverbal behavior in multiparty interaction. Rather, there is tracking of human bodies, and in the bodies, the faces are searched. The active speaker location is determined by fusing the information provided by the audio tracker and the locations of the people in the room. In different scenarios, it was shown that the multimodal approach outperforms the audio-only system.

Ba and Odobez [9] propose a novel method for head-pose-based recognition of visual focus of attention. The proposed method is based on the results from cognitive science on saccadic eye motion, which allow for the prediction of the head pose given a gaze target. More specifically, the method models head-pose observations using either a Gaussian mixture model or a hidden Markov model, whose hidden states correspond to a set of predefined visual targets. Experiments have been conducted in a meeting room, where visual targets include meeting participants, projection screen, table, and similar. The reported results clearly demonstrate that personalized models, which account for meeting-participant specific-gazing behavior, result in better recognition results.

The paper by Dai *et al.* [10] investigates the computer understanding of human group interactions in the dynamic context of a meeting scenario. To analyze these group interactions, a novel dynamic context model is developed, which coordinates bottom–up and top–down information to derive an efficient reasoning mechanism called an event-driven multilevel dynamic Bayesian network. Experiments in a "smart meeting room" demonstrate the effectiveness of this reasoning mechanism.

Tracking the behavior of groups of people and gathering interaction data are the topics of the paper by Olguin Olguin *et al.* [11]. The group's members wear a communication badge that can recognize human activity and can measure physical proximity, face-to-face interaction, and conversational time. Gathered data can be combined with digital interaction data (for example, e-mail) in order to get a more complete view of interaction patterns in an organization. In the paper, several experiments are reported, which provide information about the relation between face-to-face interaction and electronic communication and between communication patterns and job satisfaction. Knowledge about these automatically obtained behavior patterns can not only guide an organization in making changes in its physical and information technology environment but also in evaluating the effect of these changes.

The paper by Brdiczka *et al.* [12] addresses the problem of learning-situation models for context-aware services. Context is represented by a situation model consisting of the environment, users, and activities, and a framework is proposed for acquiring and evolving multilayered situations models. The approach has been evaluated in an "intelligent home" environment, and the results demonstrate the utility of this approach.

Gunes and Piccardi [13] investigate automatic recognition of a set of affective states and their temporal phases (onset, apex, and offset) from two visual cues, the face (i.e., facial expression) and the upper body (i.e., movements of hands and shoulders), in acted data. In addition to "basic" emotions, such as happiness and anger, they also explore the recognition of nonbasic emotions, such as uncertainty and puzzlement.

The paper by Doshi *et al.* [14] introduces a novel laser-based heads-up windshield display which can actively interface with human drivers. This dynamic active display presents safety-critical icons to the driver in a manner that minimizes the deviation of their gaze direction without adding unduly to the overall visual clutter. Three different display configurations have been tested and compared with a conventional display, demonstrating the displays' ability to produce very substantial improvements in driver performance.

In summary, this Special Issue attests that the research in human computing, including analysis and understanding of group interactions and context sensing and modeling, has witnessed a good deal of progress. Of course, there remain significant scientific and technical challenges to be addressed [1], [2]. However, we are optimistic about the future progress in the field. The main reason is that W5+ methodology and human computing are likely to become the single most widespread research topic in the artificial intelligence (if not of the whole computing) community. This is aided and abetted by a large and steadily growing number of research projects concerned with the interpretation of human behavior at a deeper level in group interactions [e.g., European Commission (EC) FP6 AMI/AMIDA; www.amiproject.org], any social interactions (EC FP7 Social Signal Processing NoE; www.sspnet.eu), emotionally colored HCI (e.g., EC FP7 SEMAINE; www.semaine-project.eu), etc.

ACKNOWLEDGMENT

The Guest Editors of this Special Issue would like to thank the reviewers who have volunteered their time to provide valuable feedback to the authors. They would also like to thank the contributors for making this issue an important asset to the existing body of literature in the field and also the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS editorial support for their help during the preparation of this issue.

MAJA PANTIC, *Guest Editor*
Imperial College London
Computing Department
SW7 2AZ London, U.K.
University of Twente
Electrical Engineering, Mathematics, and
Computer Science Faculty
7522 NB Enschede, The Netherlands
(e-mail: m.pantic@imperial.ac.uk)

ALEX PENTLAND, *Guest Editor*
Massachusetts Institute of Technology
Media Laboratory
Cambridge, MA 02139-4307 USA
(e-mail: pentland@media.mit.edu)

ANTON NIJHOLT, *Guest Editor*
University of Twente
Electrical Engineering, Mathematics, and
Computer Science Faculty
7522 NB Enschede, The Netherlands
(e-mail: anijholt@ewi.utwente.nl)

REFERENCES

[1] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human computing and machine understanding of human behavior: A survey," in *Artificial Intelligence for Human Computing*, vol. 4451. New York: Springer-Verlag, 2007, pp. 47–71.

[2] M. Pantic, A. Nijholt, A. Pentland, and T. Huang, "Human-Centred Intelligent Human–Computer Interaction (HCI): How far are we from attaining it?" *Int. J. Auton. Adaptive Commun. Syst.*, vol. 1, no. 2, pp. 168–187, 2008.

[3] H. Gunes, M. Piccardi, and M. Pantic, "From the lab to the real world: Affect recognition using multiple cues and modalities," in *Affective Computing: Focus on Emotion Expression, Synthesis, and Recognition*, J. Or, Ed. Vienna, Austria: I-Tech Edu. Publishing, 2008, pp. 185–218.

[4] Z. Zeng, M. Pantic, G. I. Roisman, and T. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, 2009. to be published.

[5] A. Vinciarelli, M. Pantic, H. Bourlard, and A. Pentland, "Social signal processing: Survey of an emerging domain," *Image Vis. Comput. J.*, vol. 27, 2009.

[6] D. Cook and S. Das, *Smart Environments: Technology, Protocols & Applications*. Hoboken, NJ: Wiley, 2005.

[7] L. Maat and M. Pantic, "Gaze-X: Adaptive, affective, multimodal interface for single-user office scenarios," in *Artificial Intelligence for Human Computing*, vol. 4451. New York: Springer-Verlag, 2007, pp. 251–271.

[8] F. Talantzis, A. Pnevmatikakis, and A. G. Constantinides, "Audio-visual active speaker tracking in cluttered indoors environments," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 7–15, Feb. 2009.

[9] S. O. Ba and J.-M. Odobez, "Recognizing visual focus of attention from head pose in natural meetings," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 16–33, Feb. 2009.

[10] P. Dai, H. Di, L. Dong, L. Tao, and G. Xu, "Group interaction analysis in dynamic context," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 34–42, Feb. 2009.

[11] D. Olguín Olguín, B. N. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland, "Sensible organizations: Technology and methodology for automatically measuring organizational behavior," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 43–55, Feb. 2009.

[12] O. Brdiczka, J. L. Crowley, and P. Reignier, "Learning situation models in a smart home," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 56–63, Feb. 2009.

[13] H. Gunes and M. Piccardi, "Automatic temporal segment detection and affect recognition from face and body display," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 64–84, Feb. 2009.

[14] A. Doshi, S. Y. Cheng, and M. M. Trivedi, "A novel active heads-up display for driver assistance," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 85–93, Feb. 2009.

Maja Pantic (M'98–SM'06) received the M.S. and Ph.D. degrees in computer science from the Delft University of Technology, Delft, The Netherlands, in 1997 and 2001, respectively.

She is currently a Reader in multimodal human–computer interaction (HCI) with the Computing Department, Imperial College London, London, U.K., and a Professor of affective and behavioral computing with the Electrical Engineering, Mathematics, and Computer Science Faculty, University of Twente, Enschede, The Netherlands. Her research interests include computer vision and machine learning applied to face and body gesture recognition, human communicative behavior analysis, multimodal HCI, affective computing, and e-learning tools. She published more than 90 research papers on these topics.

Dr. Pantic was the recipient of the Innovational Research Award of the Dutch Scientific Organization for her research on Facial Information for Advanced Interface in 2002, as one of the seven best young scientists in exact sciences in The Netherlands. In 2008, for her research on Machine Analysis of Human Naturalistic Behavior, she received the European Research Council Starting Grant as one of the 2.5% best young scientists in any research field in Europe. She is an Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS and of the *Image and Vision Computing Journal*. She is a Guest Editor, Organizer, and Committee member of over ten major journals and conferences in the field.

Alex Pentland (M'80–SM'99) received the B.G.S. degree from the University of Michigan, Ann Arbor, in 1976 and the Ph.D. degree from the Massachusetts Institute of Technology (MIT), Cambridge, in 1982.

He is currently with MIT, where he was the founding Director of the Media Lab Asia and currently is the Academic Head of the Media Laboratory, the Toshiba Professor of media arts and sciences, and the Director of Human Dynamics Research. His research interests include wearable computers, health systems, smart environments, and technologies for developing countries. He is a pioneer in organizational engineering, mobile information systems, and computational social science. His focus is the development of human-centered technology and the creation of ventures that take this technology into the real world.

For his work, Dr. Pentland has won numerous international awards in the arts, sciences, and engineering.

Anton Nijholt received the M.S. degree in mathematics and computer science from the Delft University of Technology, Delft, The Netherlands, and the Ph.D. degree from Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.

He was with various universities in The Netherlands, Belgium, and Canada. Currently, he is a Full Professor and the Chair of the Human Media Interaction Group, Department of Computer Science, University of Twente, Enschede, The Netherlands. His main research interests include multiparty and multimodal interaction, virtual environments, and social and intelligent (embodied) agents. He published more than 200 research papers on these topics. In addition to several large Dutch national projects, he is currently involved in research of the FP6 EC projects Augmented Multiparty Interaction and AMIDA and the FP6 EC NoE HUMAINE (the role of affect in the interface).