Björn W. Schuller, Rosalind Picard, Elisabeth André,
Jonathan Gratch, and Jianhua Tao

Intelligent Signal Processing for Affective Computing

Affective computing is computing that relates to, arises from, or deliberately influences emotion or other affective phenomena. Human emotion and affect in general are fundamental to human experience, influencing cognition, perception, and everyday tasks such as learning and communication, but are also fundamental to human health and well-being. The field of affective computing has matured over its roughly two-and-a-half decades coming closer than ever to the point of usage at scale. Affective computing is facing a plethora of different signal types—audio, video, and physiological signals, to name but the most dominant.

In everyday usage, affective computing has the potential to massively change how we interact with computing and robotic devices: they will be able to respond more appropriately to our emotions and moods, and able to show signs of empathy through mimicry, but may also use affective information for retrieval or their own creativity. Affective computing becoming truly robust also has the potential to change mental health care radically, once computing systems enable monitoring our health and well-being, detect early signs of depression, and/or identify signs of neurodevelopmental conditions (e.g., diagnosis of autism in children to burnout prevention), among others.

For most of these and many further applications, however, reliable assessment of human affect and affective behavior is key. At the same time, many of these applications require explainable decision making—not only when our health is concerned. A major breakthrough in the field—as has been the case in many related pattern-recognition problems—came with the advent and increasing usage of deep learning and other novel techniques in machine learning and artificial intelligence. Affective computing has been an early adopter of such methods, using, for example, long short-term memory recurrent neural networks already for a decade and a half. Furthermore, end-to-end learning systems ranging from the raw signal processing (e.g., audio and image processing) or shallow time–frequency representations, and more general unsupervised representation learning, are frequently encountered if not omnipresent in today's literature on affective computing. In addition, generative adversarial and transfer learning exploiting pretrained neural networks is on the rise, going as far as using image-pretrained convolutional networks for the representation of multimodal audio, visual, and autonomic physiology data. The latter is in particular triggered by the field's ever-dominating bottleneck of sufficient training data. While this has led to an impressive number of successes in

> **The field of affective computing has matured over its roughly two-and-a-half decades coming closer than ever to the point of usage at scale.**

boosting performance and reliability of existing affective computing systems, it came at the price of 1) a change in the processing of affective signals and 2) an often reduced transparency in the signal processing and decision-making parts. The lower explainability can be attributed to self-learned, generated, and transferred representations, and increasing data-injection both into the signal representation, but also signal preprocessing parts, such as source separation or signal restauration and enhancement, as typically employed when working with, for instance, low-resolution audiovisual data of human behavior.

In this special issue of *IEEE Signal Processing Magazine* (*SPM*), we survey recent advances in affective computing with a focus on intelligent signal processing. Nine articles selected from 29 submissions cover key topics including speech, music, facial, and general multimodal affective state recognition from individuals or groups, bias and fairness issues, as well as the application in mobile health. The tutorial-style articles aim to give an overview both of theory and applications. They also discuss major advances. Affective computing heavily relies on signal processing and machine intelligence, but we are certain we can raise the reader's appetite for the exciting and challenging field of affective computing, which has a lot on its plate for

signal processing and beyond in return. As a starter, this special issue first serves an introduction of the "classics" as to modalities in the recognition of (human) affective states—speech, facial expression, and physiological data.

In this issue

The first two articles deal with speech, and are to be considered as belonging together, as they separately first familiarize the reader with representation of speech in affective computing, and then with the modeling of affect based on this representation. In the first article, "On the Evolution of Speech Representations for Affective Computing: A Brief History and Critical Overview," Alisamir and Ringeval feature speech emotion recognition—an idea that was patented in the late 1970s. The authors introduce the history of the field and discuss various speech representations used over the evolution of the field. Next, in "Deep Representation Learning for Affective Speech Signal Analysis and Processing: Preventing Unwanted Signal Disparities," Lee et al. introduce the reader to the present state in speech emotion recognition—deep representation learning approaches, however, focusing mostly on the decision-making part and hence, perfectly complementing the previous article.

The issue then caters for video as input signal. "The Hitchhiker's Guide to Bias and Fairness in Facial Affective Signal Processing: Overview and Techniques," by Cheong et al., targets facial affect recognition under the particular viewpoint of fairness. The authors provide the reader with the prevailing definitions of bias and measures of fairness used in facial affective signal processing. They then categorize the algorithms and techniques for identifying and circumventing such bias. They also provide a discussion and future avenues. Also related to facial information, in "Facial-Video-Based Physiological Signal Measurement: Recent Advances and Affective Applications," Yu et al. present the reader with the

idea to assess physiological measures directly from facial information. This way, these signals, which bear crucial information on human affect, can be sensed contactless.

Expanding on single modalities, Zhao et al. then discuss the often chosen multimodal approach in "Emotion Recognition From Multiple Modalities: Fundamentals and Methodologies." In fact, a multisensorial or multimodal approach has a long tradition in affective computing, as different modalities are known for their higher relevance to certain affective states and dimensions. Hence, combining these usually allows for a more holistic and robust understanding of underlying affect. First, they introduce the usual emotion representation models and modalities used in the field. Then, they feature annotation strategies and key challenges in multimodal emotion recognition. They also show how representations can be learned per modality, how these can be fused, and how the learning algorithm can be optimized and the domain be best adapted. The reader is gently guided by real-world examples.

> **This way, these signals, which bear crucial information on human affect, can be sensed contactless.**

The articles in this special issue feature two hot topics in the field: affect of groups and fairness and bias in affective data handling. "Affects in Groups: A Review on Automated Affect Processing and Estimation in Groups" by Böck begins by providing a tutorial on group affect recognition including the latest methods and corpora in the field. The author also discusses tuning of such and next-on-the-list research for this task. In addition, "Integrating Psychometrics and Computing Perspectives on Bias and Fairness in Affective Computing: A Case Study of Automated Video Interviews" by Booth et al. features a psychometric-grounded view on bias and fairness. Broader than the previous view on these crucial aspects, it considers a more general machine learning flow. An interpersonal communication framework serves as the example of choice with mock job applications as scenario.

As an aside, this special issue offers one of the prevailing application use

cases as an example: mental health. In "Deep Learning for Mobile Mental Health: Challenges and Recent Advances," Han et al., together with some of the guest editors, provide insight into mobile device-based mental health analysis. The tutorial discusses the needed components and common approaches as well as challenges. Recent progress is discussed together with recommended future steps.

Finally, for "dessert," there will be music, as an example of media-related rather than human-related affective computing. In "Music Emotion Recognition: Toward New, Robust Standards in Personalized and Context-Sensitive Applications," Gómez-Cañón et al. introduce the task of music emotion recognition and its peculiarities such as perceived and induced emotion. They also show the typical features, data creation procedures, evaluation methods, and next steps for this field. The authors also touch on recent concerns including explainability and ethics.

Based on these insights, we come once more to the view that affective computing can be expected as a core technology not only of oncoming social robots, and general human–machine interaction, but it also has numerous applications in health, education, entertainment, security, information retrieval and recommendation systems, and many other fields that rely heavily on underlying signal processing techniques. The topic is important and timely, as 1) affective computing has matured, as mentioned, to the point of being applicable at scale, and the technology is already present in many products and 2) the landscape of methods for signal processing has significantly advanced over the last few years, with increasing data- and expert-driven intelligence injected. May this special issue help to further cross-fertilize efforts and spark ideas across affective computing and intelligent signal processing to everyone's best benefit.

Acknowledgments

Christian Jutten; Area Editor for Special Issues Xiaoxiang Zhu; and Senior Editor Meinard Müller, who guided this issue from the magazine's side, as well as *SPM* Publications Administrator Rebecca Wollman, and IEEE Magazines Managing Editor Jessica Welsh for helping with the organization and production of this issue. Most notably, we also thank the manifold reviewers for their great work during the repeated revision rounds. Without the greatest quality review, which often comes at bigger sacrifice in time, it seems impossible to reach the great quality of articles the readership expects and deserves. Finally, we thank all of the authors for serving their brief tutorialesque articles that provide the reader with the important bits and bites, while remaining light and whetting an appetite for more.

Guest Editors

Björn W. Schuller (bjoem .schuller@imperial.ac .uk) received his Ph.D. degree from the Technical University of Munich, Germany. He is a professor of artificial intelligence and head of the Group on Language and Music, Imperial College, London, SW7 2AZ, U.K., and a full professor and chair of embedded intelligence for health care and well-being at the University of Augsburg, Augsburg, 86159, Germany. He also is cofounding CEO and CSO of audEERING GmbH, providing speech emotion recognition technology. He is the field chief editor of *Frontiers in Digital Health*, former editor-in-chief of *IEEE Transactions on Affective Computing*, president emeritus of the Association for the Advancement of Affective Computing, Golden Core Awardee of the IEEE Computer Society, fellow of the British Computer Society, fellow of the International Speech Communication Association and senior member of the Association for Computing Machinery. He has authored or coauthored five books and more than 1,000 publications. He frequently consults, including Barclays, GN, Huawei, Informetis, and Samsung. He is a Fellow of IEEE.

Rosalind Picard (picard@media.mit .edu) received her B.S. degree in electrical engineering from the Georgia Institute of Technology and S.M. and Sc.D. degrees from the Massachusetts Institute of Technology (MIT), both in electrical engineering and computer science. She is a professor of media arts and sciences at MIT, founder and director of the Affective Computing Research Group at the MIT Media Lab, Cambridge, Massachusetts, 02139, USA, and cofounder of the start-ups Affectiva (acquired by Smart Eye) and Empatica. She is the author of the book *Affective Computing*, which became instrumental in helping start the field, and has published more than 350 peer-reviewed scientific articles. She has delivered more than 100 keynote speeches and regularly consults for industry. She is a Fellow of IEEE, the Association for the Advancement of Affective Computing, and a member of the SIGCHI Academy. She is a member of the National Academy of Engineering, recognized for contributions on affective computing and wearable computing.

Elisabeth André (andre@informatik .uni-augsburg.de) is a full professor of computer science and founding chair of human-centered artificial intelligence at Augsburg University, Augsburg, 86159, Germany. In 2010, she was elected a member of the Academy of Europe and the German Academy of Sciences Leopoldina. In 2017, she was elected to the CHI Academy, an honorary group of leaders in the field of human–computer interaction. To honor her achievements in bringing artificial intelligence techniques to human–computer interaction, she was awarded a EurAI fellowship in 2013. She is the editor-in-chief of *IEEE Transactions on Affective Computing*. She is a Fellow of IEEE.

Jonathan Gratch (gratch@ict.usc.edu) received his Ph.D. degree in computer science from the University of Illinois, Urbana-Champaign. He is a full research professor of computer science and psychology at the University of Southern California (USC) and director for Virtual Human Research at USC's Institute for Creative Technologies, Playa Vista, California, 90094, USA. He is the founding editor-in-chief of *IEEE Transactions on Affective Computing*, founding associate editor of *Affective Science*, associate editor of *Emotion Review* and *Journal of Autonomous Agents and Multiagent Systems*, and former president of the Association for the Advancement of Affective Computing. He is a fellow of the Association for the Advancement of Artificial Intelligence, a Cognitive Science fellow, and SIGART Autonomous Agent's Award recipient. He is a Senior Member of IEEE.

Jianhua Tao (jhtao@ nlpr.ia.ac.cn) is a professor and deputy director in the National Laboratory of Pattern Recognition, Chinese director of the Sino-European Laboratory of Informatics, Automation and Applied Mathematics, and assistant president of the Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China. His research interests include speech recognition and synthesis, human–computer interaction, affective computing, and big data analysis. He has published more than 300 papers in journals or proceedings, including *IEEE Transactions on Audio, Speech and Language Processing*, *IEEE Transactions on Affective Computing*, *IEEE Transactions on Cyberbetics*, *IEEE Transactions on Image Processing*, and *IEEE Journal of Selected Topics in Signal Processing*. He has served as the program/general cochair of various international conferences. He was a technical chair of INTERSPEECH 2020. He is a Senior Member of IEEE.

SP