Guest Editorial: Special Section on Naturalistic Affect Resources for System Building and Evaluation

Björn Schuller, *Member*, *IEEE*, Ellen Douglas-Cowie, and Anton Batliner

There has been a constant development of larger and more naturalistic databases in the field of Affective Computing over the last decade—still, the bottleneck remains: Many research issues in the field can hardly be addressed due to the evident lack of suitable data, and the existing large body of mono-modal data needs the addition of multimodal resources stemming from different modalities such as audio, video, physiology, text, etc.—in particular in real-life context and interaction: Emotional behavior is of great importance for social interaction as emotions serve communicative and social functions and convey information about people's thoughts and intentions toward others. In addition, multicultural and multilingual data is still considerably sparse. This is even truer when it comes to data in natural or working system contexts. This special section focuses on the introduction, presentation, and discussion of novel and existing mono and multimodal affective resources. Alternatively, ways to better exploit existing corpora by improved standardization and combination are needed. Steps in this direction comprise mapping schemes to overcome the peculiarities of the field—such as categorical, complex, or dimensional, and unstable annotation, and measurements to automatically assess similarity, type, and quality of resources. Also needed are new ways to establish semi-supervised processing of large resources by media tagging or ways to better bundle efforts of the community, e.g., by shared and distributed collection and annotation of data. Finally, for better exchange and comparability of reported results, partitioning and evaluation strategies will benefit from further discussion. The issues mentioned may be exemplified by novel naturalistic resources or by exploiting existing ones. Articles were invited in the area of mono and multimodal resources for research on emotion and affect.

The inspiration for this special section originated at last year's International Workshop on Emotion—Corpora for Research on Emotion and Affect—organized by Laurence Devillers, Roddy Cowie, and the guest editors. This workshop suggested the growing interest in developing corpora for research on emotion and affect and the development of a range of approaches and tools for collection and annotation. In the 15 papers accepted for and presented at the workshop, a total of 21 databases were presented covering seven languages (English (32 percent), French (26 percent), German (22 percent), Hebrew, Hungarian, Italian and Russian (5 percent each)), and two nonspeech sets. Considering the type of the spoken content, in 24 percent of the corpora the speech data to be produced were predefined, as opposed to 76 percent of the sets featuring nonconstrained speech. The nature of the emotion was acted in only 23 percent, induced in 32 percent, and natural in the majority of the sets at 45 percent, clearly reflecting the current preference of more realistic display of emotion in resources. However, studio recording still prevails—87 percent of the sets discussed being recorded in such an environment and only 13 percent in real-life surroundings (call center and medical operation room). A prevailing problem also seems to be the accessibility of data—47 percent not freely available, 24 percent available under a license, and 29 percent being freely available. Looking at the model of emotion, the ratio was 3:2 in favor of categorical versus dimensional, whereby the minimum of categories was 5 and that of dimensions 2. In addition, a 3:1 ratio was given for utilizing one of these schemata exclusively versus using both. The trend behind these figures illustrates the increasing popularity of dimensional or more complex modeling. The according number of annotators in these sets varied from only 1 (in 7 percent of the sets) to 17, leading to a mean of 4 annotators (2 and 3 to 6 in 37 percent of cases, and more than 10 in 19 percent of the cases). At the same time, the number of subjects recorded spanned from 4 to 100, with a total of 530 and a mean of 38, whereby the female to male ratio was at 4:3. In terms of size, the recorded speech time varied between 47 minutes and 22 hours, leading to a total of 69 hours and a mean of 6 hours per corpus with a minimum of 120 to 13,731 instances per set and 3,641 on average. If one judged exclusively from these databases, last year's average emotion corpus would be spoken English, verbally nonrestricted, natural in terms of emotion, though studio recorded, labeled in 15 classes or 4 dimensions by 4 annotators, contain 38 speakers with more of them being females, and resulting in 6 hours of speech with 3,6 k instances. Yet, it would be proprietary and not feature a defined partitioning. This summary of the workshop points to a number of trends: More and more languages are covered; more natural databases labeled in more complex models might be available to the community in the near future. At the same time, increasingly more multimodal resources are to be expected. Finally, future initiatives could

- *B. Schuller is with the Institute for Human-Machine Communication, Technische Universität München (TUM), D-80333 München, Germany. E-amil: schuller@ieee.org.*
- *E. Douglas-Cowie is with the School of English, Queen's University, Belfast, BT7 1NN, UK. E-mail: E.Douglas-Cowie@qub.ac.uk.*
- *A. Batliner is with the Pattern Recognition Lab, Friedrich-Alexander University, Erlangen-Nuremberg, Germany. E-mail: batliner@informatik.uni-erlangen.de.*

help foster combined community efforts for merging and common labeling of resources, and make such desperately needed larger amounts of resources accessible.

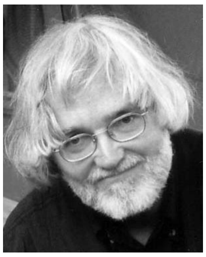
For this special section, we received 13 articles, of which four were carefully selected (i.e., 30 percent acceptance rate) as follows: The first article, "The SEMAINE Database: Annotated Multimodal Records of Emotionally Colored Conversations between a Person and a Limited Agent" by Gary McKeown, Michel Valstar, Roddy Cowie, Maja Pantic, and Marc Schröder, deals with high quality audiovisual recordings of 150 participants interacting with different configurations of a Sensitive Artificial Listener (SAL) agent, for 959 conversations (approx. 5 min. each); 6-8 raters and 27 associated categories were employed. Next, "DEAP: A Database for Emotion Analysis using Physiological Signals" by Sander Koelstra, Christian Mühl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras, describes a multimodal data set with physiological and (partly) video recordings, with 32 participants, watching and rating music videos; the data are analyzed and correlates between signals and ratings as well as classification performance for arousal and valence are reported. Then, in "The Belfast Induced Natural Emotion Database" by Ian Sneddon, Margaret McRorie, Gary McKeown, and Jennifer Hanratty, the authors describe recordings of mild to moderate emotionally colored responses to a series of laboratory-based emotion induction tasks; self-reports, continuous trace-style ratings of dimensions, and several other characterizing parameters are available. The last article, "A Multimodal Affective Database for Affect Recognition and Implicit Tagging" by Mohammad Soleymani, Jeroen Lichtenauer, and Maja Pantic, introduces MAHNOB-HCI, a multimodal affective database: 30 participants watched 20 emotional videos and self-reported their felt emotions and judged videos/images with/without correct/incorrect emotion tags. Recognition results are discussed for thee classes and two dimensions.

ACKNOWLEDGMENTS

We would like to thank the editor in chief, Jonathan Gratch, for his help with this special issue and the 35 reviewers that not only helped with the decision process but contributed with excellent reviews to make this issue special: Alexandra Balahur, Nadia Berthouze, Yves Bestgen, Felix Burkhardt, Carlos Busso, Ginevra Castellano, Jeffrey Cohn, Roddy Cowie, Laurence Devillers, Florian Eyben, Anna Esposito, Christos Frantzidis, Roland Goecke, Hatice Gunes, Dirk Heylen, Diana Inkpen, Kostas Karpouzis, Abe Kazemzadeh, Jonghwa Kim, Kolja Kühnlenz, Bruno Lepri, Jean-Claude Martin, Marie-Francine Moens, Emily Mower, Alexander Osherenko, Massimo Piccardi, Michael Popp, Tanja Schultz, Dino Seppi, Stefan Steidl, Alessandro Valitutti, Michel Valstar, Nigel Ward, Felix Weninger, and Martin Wöllmer.

Björn Schuller received the diploma in 1999 and the doctoral degree in 2006, both in electrical engineering and information technology from the Technische Universität München (TUM), where he has since been tenured as a senior researcher and lecturer in pattern recognition and speech processing. From 2009 to 2010 he was with the CNRS-LIMSI Spoken Language Processing Group in Orsay, France, and a visiting scientist at Imperial College London's Department of Computing in London, United Kingdom. He is a member of the ACM, HUMAINE Association, IEEE, and ISCA and has (co)authored two books and more than 200 peer reviewed publications, leading to more than 2,400 citations—his current H-index equals 26. He serves as a member and secretary of the steering committee and as an associate and guest editor of the *IEEE Transactions on Affective Computing*, as a guest editor and reviewer for more than 30 leading journals and multiple conferences in the field, and as an invited speaker, session and challenge organizer, including the INTERSPEECH 2009 Emotion, 2010 Paralinguistic, and 2011 Speaker State Challenges and chairman and program committee member of numerous international workshops and conferences.

Ellen Douglas-Cowie graduated with the BA and DPhil degrees in sociolinguistics from the New University of Ulster (1972). She became a lecturer in linguistics at Queen's University Belfast in 1975, and a professor in 2003, since when she has been Dean of Faculty and pro-Vice-Chancellor. Her research has studied the characteristics that distinguish varieties of speech—clinical, social, and stylistic—and includes seminal papers on sociolinguistics and deafened speech. She has particular expertise in the analysis of prosody and in fieldwork, particularly the collection of data. She has worked on a series of projects concerned with the identification and recognition of emotion from speech and face (see above), and led the HUMAINE work package on databases. High profile output includes coeditorship of an influential special edition of *Speech Communication* (2003).



Anton Batliner received the MA degree in Scandinavian languages and the Drphil degree in phonetics in 1978, both from LMU Munich. He has been a member of the research staff of the Institute for Pattern Recognition at Friedrich-Alexander University (FAU) since 1997. He is coeditor of one book and author/coauthor of more than 200 technical articles, with a current H-index of 30 and more than 3,500 citations. His research interests are the modeling and automatic recognition of emotional user states, all aspects of prosody and paralinguistics in speech processing, uni and multimodal focus of attention, pronunciation assessment, and spontaneous speech phenomena such as disfluencies, irregular phonation, etc. He served as Workshop/Session (co)organizer for Emotional Corpora I, II, III (LREC), Paralinguistics (ICPhS 07), Non-prototypical Emotions (ACCI 09), Emotion Challenge (INTERSPEECH 2009), Paralinguistic Challenge (INTERSPEECH 2010), Computer Aided Pronunciation Training (Prosody 2010); he was a guest editor for *AHCI*, *Computer Speech and Language*, and *Speech Communication*, and is associate editor for the *IEEE Transactions on Affective Computing* as well as a reviewer for numerous leading journals and conferences.