Preface

The research in Human-Machine Interaction (HMI) is an emerging topic in various research communities. It is not only a matter in computer sciences but moreover, it is a subject of psychology, engineering, neuroscience, cognitive science, and many related disciplines. The connecting question for all the researchers in the above-mentioned communities is:

> How can the user's behavior be analyzed to improve the interaction between humans and machines?

For this, we usually investigate at first the interaction between humans to understand how the fragile interplay and interrelationship of communication partners is established and afterwards pursued. The complex scenario of a dialog between (even two) interlocutors is enriched with subtle reactions and characteristics. Just for human observers, a correct judgement according to aspects like social context and background, emotions and feelings, possible reactions, is quite challenging. Fortunately, a human observer is able to incorporate a huge variety of multiple knowledge sources and multimodal sensor inputs. Such knowledge is, for instance, the contextual information of the current interaction, the social and cultural background of the communication partners, or the possible and intended goal of the discussion. On the other hand, sensory inputs like speech and sound, gaze, facial expression, gestures can be used to enhance and enable a detailed analysis of the interaction process.

In recent research activities, the aim is to transfer such considerations to HMI. Like a human observer, a machine should use as much information as possible to derive a correct judgement of an interaction. A multimodal investigation of the communication is, therefore, mandatory in order to understand valuable cues that have to be considered in the machine's internal situation evaluation. Such advice applies not only to the analysis part but also to the output or interface design. In general, a suitable interface will consider both input and output in an appropriate way. For this, one of the goals in building multimodal user interfaces is to make the interaction between user and system as natural as possible. Again, the most natural form of interaction is how humans interact with each other.

While the analysis of human-human communication and HMI has resulted in many insights and further, an increased understanding of interaction, transferring this knowledge to a real-time situation, is still challenging. To use technical systems in a proper way in daily interaction a real-time evaluation is necessary and important. It requires that input from a user, coming from speech, gaze, facial expression, and any other modality is recorded and interpreted in real-time or with minimal delay. The interpretation can be either semantic, or could aim at more affective properties such as the personality, mood, or intentions of the user. For this, the combined effort of multiple aforementioned disciplines have to be taken into account to achieve such a goal. Finally, a system or an agent also needs to respond appropriately to the user without delays to ensure that the interaction is smooth.

To establish systems or, in a sense of a more human-like interpretation, (artificial virtual) agents, we identify three main topics to be emphasized as important research areas.

Multimodal Annotation. The generation of agents and systems is based on data. In general, most communities prefer datasets that are already preprocessed and thus, suitable for direct development. On the other hand, this implies that the corpora contain significant material which will be usually explored and provided during annotation. In closed connection also issues of proper features to be applied in systems are important to discuss. Notice that such a processing of data is not only focused on annotation but covers all steps from data collection, (pre-)processing, annotation up to feature extraction and (sub-)symbolic interpretation of the given material. As we already discussed, a proper investigation of an interaction is mostly valuable in a multimodal fashion. Hence, existing approaches have to be adapted and – if necessary – enhanced to be feasible for further analyses. Currently, the idea of open data is an emerging topic and thus, should be linked towards multimodal data.

Multimodal Analyses. The analyses of data are complex. Each observed modality has its own characteristics which have to be considered in detail. For instance, from speech various aspects could be covered: the contextual information, the prosodic and para-linguistic features, and intentional cues. The same variety can be seen in video material where the analysis of several signals, such as eye gaze, gestures, postures, is a key element. Real-time processing of all this is the real challenge. In this context, questions of fusion as well as combination of features and results are discussed in recent research.

Applications and Systems. Finally, the achievements of analyses will result in systems and (artificial virtual) agents combining proper inputs and outputs to handle an interaction in a human-like manner. The development of feasible applications is a crucial and challenging issue. For this, generally, user studies are conducted which are based on Wizard-of-Oz or mock-up scenarios. Nevertheless, such studies provide insights and understandings of theoretical and methodological approaches which can be directly transferred to setups of agents. Further, they usually lead to novel applications. Therefore, the development of systems and agents has to be fostered.

Based on these considerations and having in mind the challenges of current HMI and (artificial virtual) agents, we conceptualized the 2nd International Workshop on Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction (MA3HMI). The workshop was held in conjunction with INTERSPEECH 2014, on September 14, 2014 in Singapore.

The MA3HMI workshop aimed to bring together researchers working on the analysis of multimodal recordings as a means to develop systems that can interact with humans. (Artificial) agents can be regarded in their broadest sense, including virtual chat agents, empathic speech interfaces, and life-style coaches on a smart phone. Complementary to the 2012 edition of MA3HMI, the focus of the 2014 MA3HMI was on speech which transfers both content and social information. We were particularly interested in speech technologies for HMI, and the combination of speech and natural language processing with the analysis of other modalities.

We encouraged researchers to present and discuss their papers that concern the different development phases of HMI, including the recording and online analysis of multimodal conversations, the modeling of the dialog, and the user evaluation of such systems. Further, tools and systems that address real-time conversations with (artificial virtual) agents were also within the topics of MA3HMI.

From the submitted contributions, which received at least two independent reviews, nine papers were selected for oral presentation and publication in this issue of Lecture Notes on Artificial Intelligence (LNAI). The papers were grouped in the sections "Human-Machine Interaction" and "Dialogs and Speech Recognition" which serve also as main parts of this book.

In addition to the oral presentations, we had a lively and really interesting plenary discussion on hot topics in the context of HMI and agents. We thank all authors and participants of the MA3HMI workshop for their contribution.

Furthermore, the workshop was enriched by an extraordinary keynote talk given by Prof. Nick Campbell, Speech Communication Lab at Trinity College Dublin. In his plenary talk Prof. Campbell presented a novel corpus for multimodal studies of interactions. Further, he provided insights into the data generation and processing of such multimodal datasets. Moreover, Nick Campbell shared his ideas and thoughts on analyses and what to do and where to go in the research of HMI. We thank Prof. Nick Campbell for accepting our invitation for a keynote talk, his contribution, and the subsequent discussion.

Further, we thank the members of the Program Committee for their effort in reviewing the submissions and identifying the most relevant papers for the year 2014 MA3HMI.

Our sincere gratitude goes also to Springer and to Alfred Hofmann and Anna Kramer as well as their team for their continuous support and all the effort in preparing the current issue of LNAI.

We thank the INTERSPEECH 2014 workshops chair, Chai Wutiwiwatchai, and the local arrangement team for their support. Without their help the workshop's venue would not have been so well prepared. Further, we also acknowledge Singapore Expo and their collaborators for hosting the workshop.

Finally, our gratitude goes to our generous sponsors, the Transregional Collaborative Research Center "Companion Technology" and the FastNet project, which provided financial support for MA3HMI. Further, we acknowledge the endorsement of ISCA.

October 2014

Ronald Böck
Francesca Bonin
Nick Campbell
Ronald Poppe

http://www.springer.com/978-3-319-15556-2

Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction
Second International Workshop, MA3HMI 2014, Held in Conjunction with INTERSPEECH 2014, Singapore, Singapore, September 14, 2014, Revised Selected Papers
Böck, R.; Bonin, F.; Campbell, N.; Poppe, R. (Eds.)
2015, XII, 109 p. 29 illus., Softcover
ISBN: 978-3-319-15556-2