Nonlinear Systems for multimodal Human-Machine Interaction

Anna Esposito and Stefano Squartini*

COST 2102 Special Session Proposal for WIRN2010

I. THEME AND SCOPE OF THE SESSION

COST (European Cooperation in the Field of Scientific and Technical Research) is a widely-used tool by European Commission to support co-operation among scientists and researchers across Europe. Several different Actions have been funded in the last decade. COST2102 is an ICT (Information and Communication Technology) Action, titled as Cross-Modal Analysis of Verbal and Non-Verbal Communication. As mentioned in related MoU (Memorandum of Understanding) "the main objective of the Action is to develop an advanced acoustical, perceptual and psychological analysis of verbal and non-verbal communication signals originating in spontaneous face-to-face interaction, in order to identify algorithms and automatic procedures capable of identifying human emotional states. Several key aspects will be considered, such as the integration of the developed algorithms and procedures for application in telecommunication, and for the recognition of emotional states, gestures, speech and facial expressions, in anticipation of the implementation of intelligent avatars and interactive dialogue systems that could be exploited to improve user access to future telecommunication services.”

Nonlinear Systems have shown great capabilities in modeling, prediction, recognition tasks in Digital Signal Processing and a mature degree of understanding has been achieved in many application areas, such as Human-Machine Interaction (HMI), widely addressed by the literature and with a strong commercial impact also. Algorithms and systems developed in this area typically involve the usage of a single-type of communication channel, even though a certain attention has been recently paid to the exploitation of signal multi-modality (including the human emotional states) to augment the completeness of available information and so enhance the HMI systems potentials. It is indeed of great interest for the scientific community understanding how such multi-modality can be efficiently used by the intelligent nonlinear systems devoted to solve the HMI tasks, guaranteeing the naturalness of interaction as best as possible. The attempt of merging verbal and non-verbal communication, specifically addressed by COST2102, is a significative confirm of this trend and pushed the organizers of this special session to find within WIRN2010 a fruitful scientific scenario to deepen such issues. The aim of this session is discussing the most recent technological efforts from this perspective, therefore with special focus on the employment of nonlinear solutions to solve tasks oriented to multimodal HMI applications.

II. TOPICS

Topics of the session include but are not limited to:
- Artificial Intelligence for Human-Machine Interface
- Cognitive Systems for Multimodal Signal Analysis
- Analysis and Identification of human emotional states
- Human behavior and unsupervised interactive interfaces
- Analysis of verbal and non-verbal communication signals
- Nonlinear Audio-Visual Processing
- Speech-interfaced systems for Human-Machine Interaction
- Multimodal Signal Processing for Biometric Applications

Anna Esposito is with International Institute for Advanced Scientific Studies (IIASS), Via G. Pellegrino 19 84019 - Vietri sul Mare (SA), Italy.
S.Squartini is with the 3MediaLabs, DIBET, Università Politecnica delle Marche, Via Brecce Bianche 31, 60131 Ancona, Italy.
*Corresponding author, email address: s.squartini@univpm.it, phone/fax: +39 071 220 4381/4453.
III. INVITED SPEAKER

Prof. Gerhard Rigoll, rigoll@tum.de

Institute for Human-Machine Communication, TU München, Arcisstr. 21, D-80333 Munich, Germany

Short Bio: Prof. Dr.-Ing. habil. Gerhard Rigoll obtained the Dipl.-Ing degree from Stuttgart University/ Germany, in 1982. He joined Fraunhofer-Institute (IAO) in Stuttgart and received the Dr.-Ing. degree in 1986 in the area of automatic speech recognition. From 1986 to 1988 he worked as postdoctoral fellow at IBM T.J. Watson Research Centre in Yorktown Heights/USA for the IBM Tangora speech recognition project. He received the Dr.-Ing. habil. degree in 1991 from Stuttgart University with a thesis on speech synthesis. From 1991 to 1993 he worked as a guest researcher in the framework of the EC Scientific Training Programme in Japan for the NTT Human Interface Laboratories in Tokyo/Japan. In 1993 he was appointed full professor of computer science at Gerhard-Mercator-University in Duisburg, Germany. In 2002, he joined Technische Universität München (TUM), where he is now heading the institute for Human-Machine Communication.

His research interests are in the field of human-machine communication and multimedia information processing, covering areas such as speech and handwriting recognition, gesture recognition, face detection and identification, emotion recognition, person tracking, information retrieval, video-indexing and interactive computer graphics. Dr. Rigoll is a Senior Member of the IEEE and is the author and co-author of about 400 papers in the field of pattern recognition, covering the above mentioned application areas. He was Associate Editor of the IEEE Transactions on Audio, Speech and Language Processing and other international journals, and has served as session chairman, organizer and member of the program committee for numerous international conferences.

Talk: Multimodal Interaction: Methods and Applications for Joint Cooperation Between Humans and Cognitive Systems

We present an overview on recent advances in our work on multimodal human-machine communication between humans and technical cognitive systems in order to enable the solution of complex problems that require the joint cooperation of humans and intelligent systems to accomplish challenging tasks that can be only successfully handled if both interact in a cooperative manner. The technical system has cognitive capabilities that support an intelligent interaction with the human operator, as e.g. in case of a joint assembly task of a human assisted by a cognitive robot during the manufacturing process of a complex product. The talk will start with the presentation of the major paradigms used for interaction, coming mainly from statistical pattern recognition, such as e.g. Hidden-Markov-Models and the more powerful Graphical Models. The second part of the talk will then describe the challenges in joint cooperation with cognitive systems, which occur often in intelligent manufacturing environments that provide a very rich application scenario for the successful combination of advanced multimodal human-machine communication techniques together with cognitive components for improved, cooperative interaction between technical systems and human operators.

IV. LIST OF PAPERS

1) S.Squartini, M.Fagiani, E.Principi, F.Piazza. “Multichannel Cepstral Domain Feature Warping for Robust Speech Recognition”. All authors are with 3MediaLabs, Department of Biomedical Engineering, Electronics and Telecommunications, Università Politecnica delle Marche, Ancona, Italy.
2) Maurice Grinberg, Vladimir Haltakov, Hristo Stefanov. “Spreading Activation Machanisms for Efficient Knowledge Retrieval form Large Datasets”. New Bulgarian University, Montevideo str. 21, 1618 Sofia, Bulgaria.
4) Josep Roure and Daniel Rodriguez. “Demonstration based Policy Learning in a Reduced Driving Environment”. All authors are with Escola Universitaria Politecnica de Mataró, Barcelona, Spain.
5) Anna Esposito and Maria Teresa Riviello. “On the perception of facial emotional expression in children”. All authors are Seconda Università di Napoli and IIASS, Italy.
6) Evgenia Hristova “Biosignals and decision making”. Evgenia Hristova is with New Bulgarian University, Montevideo str. 21, 1618 Sofia, Bulgaria.