Lecture Notes in Artificial Intelligence 5641

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Anna Esposito Robert Vích (Eds.)

Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions

COST Action 2102 International Conference
Prague, Czech Republic, October 15-18, 2008
Revised Selected and Invited Papers

Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Anna Esposito
Second University of Naples, Department of Psychology
and IIASS, International Institute for Advanced Scientific Studies
Via G. Pellegrino 19, 84019 Vietri sul Mare (SA), Italy
E-mail: iiass.annaesp@tin.it

Robert Vích
Institute of Photonics and Electronics
Academy of Sciences of the Czech Republic
Chaberská 57, 182 52 Prague 8, Czech Republic
E-mail: vich@ufe.cz

Library of Congress Control Number: 2009931057

CR Subject Classification (1998): I.5, H.5, I.2.7, I.2.10, I.4

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-642-03319-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-03319-3 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2009
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12731275 06/3180 5 4 3 2 1 0

This book is dedicated to:

Maria Marinaro
A person of exceptional human and ethical qualities
and a scientist of outstanding value.

and to all those:

who posit questions whose answers raise new questions
driving the emotional sense of any scientific work

Preface

This volume brings together the peer-reviewed contributions of the participants at the COST 2102 International Conference on "Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions" held in Prague, Czech Republic, October 15–18, 2008.

The conference was sponsored by COST (European Cooperation in the Field of Scientific and Technical Research, www.cost.esf.org/domains_actions/ict) in the domain of Information and Communication Technologies (ICT) for disseminating the research advances developed within COST Action 2102: "Cross-Modal Analysis of Verbal and Nonverbal Communication" http://cost2102.cs.stir.ac.uk.

COST 2102 research networking has contributed to modifying the conventional theoretical approach to the cross-modal analysis of verbal and nonverbal communication changing the concept of *face to face communication* with that of *body to body communication* as well as developing the idea *of embodied information.* Information is no longer the result of a difference in perception and is no longer measured in terms of quantity of stimuli, since the research developed in COST 2102 has proved that human information processing is a nonlinear process that cannot be seen as the sum of the numerous pieces of information available. Considering simply the pieces of information available, results in a model of the receiver as a mere decoder, and produces a huge simplification of the communication process. What has emerged from COST 2102 research is that human information processing does rely on *several communication modes* but also, more importantly, strongly depends on the context in which the communication process is instantiated. The implications are a change of perspective where the research focus moves from "communicative tools" to "communicative instances" and ask for investigations that take into account the environment and the context in which communicative acts take place. The consequences in ICT research should lead to the development of instantiated interactive dialogue systems and instantiated intelligent avatars able to act by exploiting contextual and environmental signals and to process them by combining previous experience (memory) adapted to the problem instance.

Currently, advances in COST 2102 research have shown the implementation of interactive systems such as:

- Visitors controlled an avatar in a lively multi-user 3D environment where characters follow natural and realistic behavior patterns.
- Virtual trainer that monitors the user's behavior.
- Demonstration of the use of motion capture, physical simulation and kinematics on a single body.
- Multimodal signal processing for dance synthesis by analysis.
- Alternative augmentative communication systems.
- Infant voice-controlled robot.

- Data of interactions and vocal and facial emotional expressions that are exploited for developing new algorithms and mathematical models for vocal and facial expression recognition.
- Software showing progress made in HMI, as far as spoken dialogue is concerned.
- Remote health monitoring.
- Telecommunication.

Some of these implementations and results were presented at ICT 2008 in Lyon, France, during November 23–25, 2008
http://www.cost.esf.org/about_cost/cost_stories/ICT-2008
http://www.cost.esf.org/events/ICT-2008-I-s-to-the-Future

The conference in Prague was developed around the COST 2102 main themes and benefited from a special session on "Emotions and ICT" jointly organized with COST 298 (http://www.cost298.org/).

This book is roughly arranged into three sections, according to a thematic classification, even though the COST 2102 research field is largely interdisciplinary and complex, involving expertise in computer graphics, animation, artificial intelligence, natural language processing, cognitive and psychological modeling of human–human and human–machine interaction, linguistics, communication, and artificial life and cross-fertilization between social sciences and engineering (psychology, sociology, linguistic, neuropsychology).

The first section "Emotion and ICT," deals with themes related to the cross-fertilization between studies on ICT practices of use and cross-modal analysis of verbal and nonverbal communication.

The second section, "Verbal and Nonverbal Features of Computational Phonetics," presents original studies devoted to the modeling of verbal and nonverbal phonetics.

The third section, "Algorithmic and Theoretical Analysis of Multimodal Interfaces," presents theoretical and practical implementations of original studies devoted to the analysis of speech, gestures, face and head movements as well as to learning issues in human–computer interaction and to algorithmic solutions for noise environments in human–machine exchanges.

The editors would like to thank the COST- ICT Programme for supporting the realization of the conference and the publication of this volume, and in particular the COST Science Officers Gian Mario Maggio, Francesca Boscolo and Sophie Beaubron for their constant help, and guidance.

Our gratitude goes to the staff of the Charles University in Prague, and in particular to Jan Volin for making available the space and people to help in the conference organization. The Prague Academia of Science, Institute of Photonics and Electronics, is deeply acknowledged for contributing to the event, in particular, Petr Horák for his hard and invaluable work.

Special appreciation goes to the International Institute for Advanced Scientific Studies, and in particular to Tina Marcella Nappi, Michele Donnarumma, and Antonio Natale, for their invaluable editorial and technical support in the organization of this volume .

The editors are extremely grateful to the contributors and the keynote speakers, whose work stimulated an extremely interesting interaction with the attendees, and to COST 2102 International Scientific Committee for the accurate review work, for their dedication, and their valuable selection process.

May 2009

Anna Esposito
Robert Vích

Organization

International Advisory and Organizing Committee

Robert Vích	Institute of Photonics and Electronics, Prague, Czech Republic
Anna Esposito	Second University of Naples and IIASS, Italy
Eric Keller	University of Lausanne, Switzerland
Macos Faundez-Zanuy	University of Mataro, Barcelona, Spain
Petr Horák	Institute of Photonics and Electronics, Prague, Czech Republic
Amir Hussain	University of Stirling, UK
Dagmar Dvořáková	Media Communication Department, Prague, Czech Republic
Jitka Veroňková	Institute of Phonetics, Charles University, Prague, Czech Republic
Jitka Pečenková	Institute of Photonics and Electronics, Prague, Czech Republic
Irena Vítková	Media Communication Department, Prague, Czech Republic
Jan Volín	Institute of Phonetics, Charles University, Prague, Czech Republic

International Scientific Committee

Uwe Altmann	Technische Universität Dresden, Germany
Hicham Atassi	Brno University of Technology, Czech Republic
Nikos Avouris	University of Patras, Greece
Ruth Bahr	University of South Florida, USA
Gérard Bailly	ICP, Grenoble, France
Marian Bartlett	University of California, San Diego, USA
Štefan Beňuš	Constantine the Philosopher University, Nitra, Slovakia
Niels Ole Bernsen	University of Southern Denmark, Denmark
Jonas Beskow	Royal Institute of Technology, Sweden
Horst Bishof	Technical University Graz, Austria
Peter Birkholz	Aachen University, Germany
Jean-Francois Bonastre	Universitè d'Avignon, France
Nikolaos Bourbakis	ITRI, Wright State University, Dayton, USA
Maja Bratanić	University of Zagreb, Croatia
Antonio Calabrese	Istituto di Cibernetica – CNR, Naples, Italy
Paola Campadelli	Università di Milano, Italy
Nick Campbell	ATR Human Information Science Labs, Kyoto, Japan
Antonio Castro Fonseca	Universidade de Coimbra, Portugal

Aleksandra Cerekovic	Faculty of Electrical Engineering, Croatia
Josef Chaloupka	Technical University of Liberec, Czech Republic
Mohamed Chetouani	Universitè Pierre et Marie Curie, France
Gerard Chollet	CNRS-LTCI, Paris, France
Muzeyyen Ciyiltepe	Gulhane Askeri Tip Academisi, Ankara, Turkey
Anton Čižmár	Technical University of Košice, Slovakia
Nicholas Costen	Manchester Metropolitan University, UK
Vlado Delic	University of Novi Sad, Serbia
Marion Dohen	ICP, Grenoble, France
Francesca D'Olimpio	Second University of Naples, Italy
Thierry Dutoit	Faculté Polytechnique de Mons, Belgium
Laila Dybkjær	University of Southern Denmark, Denmark
Matthias Eichner	Technische Universität Dresden, Germany
Aly El-Bahrawy	Faculty of Engineering, Cairo, Egypt
Engin Erzin	Koc University, Istanbul, Turkey
Anna Esposito	Second University of Naples, and IIASS, Italy
Joan Fàbregas Peinado	Escola Universitaria de Mataro, Spain
Sascha Fagel	Technische Universität Berlin, Germany
Nikos Fakotakis	University of Patras, Greece
Marcos Faundez-Zanuy	Escola Universitaria de Mataro, Spain
Dilek Fidan	Ankara University, Turkey
Leopoldina Fortunati	Università di Udine, Italy
Carmen García-Mateo	University of Vigo, Spain
Björn Granström	Royal Institute of Technology (KTH), Sweden
Marco Grassi	Università Politecnica delle Marche, Italy
Maurice Grinberg	New Bulgarian University, Bulgaria
Mohand Said Hacid	Universitè Claude Bernard Lyon 1, France
Jaakko Hakulinen	University of Tampere, Finland
Ioannis Hatzilygeroudis	University of Patras, Greece
Immaculada Hernaez	University of the Basque Country, Spain
Javier Hernando	Technical University of Catalonia, Spain
Wolfgang Hess	Universität Bonn, Germany
Dirk Heylen	University of Twente, The Netherlands
Rüdiger Hoffmann	Technische Universität Dresden, Germany
David House	Royal Institute of Technology (KTH), Sweden
Amir Hussain	University of Stirling, UK
Ewa Jarmolowicz	Adam Mickiewicz University, Poznan, Poland
Kristiina Jokinen	University of Helsinki, Finland
Jozef Juhár	Technical University Košice, Slovak Republic
Zdravko Kacic	University of Maribor, Slovenia
Maciej Karpinski	Adam Mickiewicz University, Poznan, Poland
Eric Keller	Université de Lausanne, Switzerland
Adam Kendon	University of Pennsylvania, USA
Stefan Kopp	University of Bielefeld, Germany
Jacques Koreman	University of Science and Technology, Norway
Robert Krauss	Columbia University, New York, USA
Maria Koutsombogera	Inst. for Language and Speech Processing, Greece

Bernd Kröger	Aachen University, Germany
Gernot Kubin	Graz University of Technology, Austria
Alida Labella	Second University of Naples, Italy
Yiannis Laouris	Cyprus Neuroscience and Technology Institute, Cyprus
Børge Lindberg	Aalborg University, Denmark
Wojciech Majewski	Wroclaw University of Technology, Poland
Pantelis Makris	Neuroscience and Technology Institute, Cyprus
Raffaele Martone	Second University of Naples, Italy
Dominic Massaro	University of California - Santa Cruz, USA
David McNeill	University of Chicago, USA
Nicola Melone	Second University of Naples, Italy
Katya Mihaylova	University of National and World Economy, Sofia, Bulgaria
Michal Mirilovič	Technical University of Košice, Slovakia
Peter Murphy	University of Limerick, Ireland
Antonio Natale	Salerno University and IIASS, Italy
Eva Navas	Escuela Superior de Ingenieros, Bilbao, Spain
Delroy Nelson	University College London, UK
Géza Németh	Budapest University of Technology, Hungary
Friedrich Neubarth	Research Inst. Artificial Intelligence, Austria
Giovanna Nigro	Second University of Naples, Italy
Anton Nijholt	University of Twente, The Netherlands
Jan Nouza	Technical University of Liberec, Czech Republic
Igor Pandzic	Faculty of Electrical Engineering, Croatia
Harris Papageorgiou	Inst. for Language and Speech Processing, Greece
Ana Pavia	Spoken Language Systems Laboratory, Portugal
Catherine Pelachaud	Université de Paris 8, France
Bojan Petek	University of Ljubljana, Slovenia
Harmut R. Pfitzinger	University of Munich, Germany
Francesco Piazza	Università Politecnica delle Marche, Italy
Neda Pintaric	University of Zagreb, Croatia
Isabella Poggi	Università di Roma 3, Italy
Jiří Přibil	Academy of Sciences, Czech Republic
Anna Přibilová	Slovak University of Technology, Slovakia
Michael Pucher	Telecommunications Research Center Vienna, Austria
Jurate Puniene	Kaunas University of Technology, Lithuania
Giuliana Ramella	Istituto di Cibernetica – CNR, Naples, Italy
Kari-Jouko Räihä	University of Tampere, Finland
José Rebelo	Universidade de Coimbra, Portugal
Luigi Maria Ricciardi	Università di Napoli "Federico II", Italy
Matej Rojc	University of Maribor, Slovenia
Algimantas Rudzionis	Kaunas University of Technology, Lithuania
Vytautas Rudzionis	Kaunas University of Technology, Lithuania
Milan Rusko	Slovak Academy of Sciences, Slovak Republic
Zsófia Ruttkay	Pazmany Peter Catholic University, Hungary
Bartolomeo Sapio	Fondazione Ugo Bordoni, Rome, Italy
Yoshinori Sagisaka	Waseda University, Tokyo, Japan

Silvia Scarpetta	Salerno University, Italy
Ralph Schnitker	Aachen University, Germany
Jean Schoentgen	Université Libre de Bruxelles, Belgium
Stefanie Shattuck-Hufnagel	MIT, Cambridge, USA
Zdeněk Smékal	Brno University of Technology, Czech Republic
Stefano Squartini	Università Politecnica delle Marche, Italy
Piotr Staroniewicz	Wroclaw University of Technology, Poland
Vojtěch Stejskal	Brno University of Technology, Czech Republic
Marian Stewart-Bartlett	University of California, San Diego, USA
Jianhua Tao	Chinese Academy of Sciences, P.R. China
Jure F. Tasič	University of Ljubljana, Slovenia
Murat Tekalp	Koc University, Istanbul, Turkey
Kristinn Thórisson	Reykjavík University, Iceland
Isabel Trancoso	Spoken Language Systems Laboratory, Portugal
Luigi Trojano	Second University of Naples, Italy
Wolfgang Tschacher	University of Bern, Switzerland
Markku Turunen	University of Tampere, Finland
Henk Van Den Heuvel	Radboud University Nijmegen,The Netherlands
Robert Vích	Academy of Sciences, Czech Republic
Klára Vicsi	Budapest University of Technology, Hungary
Leticia Vicente-Rasoamalala	Alchi Prefectural Univesity, Japan
Hannes Högni Vilhjálmsson	Reykjavík University, Iceland
Jane Vincent	University of Surrey, Guilford, UK
Vogel	University of Dublin, Ireland
Jan Volín	Charles University, Czech Republic
Rosa Volpe	Université De Perpignan Via Domitia, France
Yorick Wilks	University of Sheffield, UK
Matthias Wimmer	Technische Universiät München, Germany
Matthias Wolf	Technische Universität Dresden, Germany
Bencie Woll	University College London, UK
Bayya Yegnanarayana	Institute of Information Technology, India
Jerneja Žganec Gros	Alpineon Development and Research, Slovenia
Goranka Zoric	Faculty of Electrical Engineering, Croatia

Sponsors

- COST - European Science Foundation: COST ACTION 2102: “Cross-Modal Analysis of Verbal and Nonverbal Communication”
- Institute of Photonics and Electronics, Academy of Sciences, Prague, Czech Republic
- Media and Communication, Academy of Sciences, Prague, Czech Republic
- Institute of Phonetics, Faculty of Philosophy and Arts, Charles University, Prague, Czech Republic

- Institute of Applied Physics, Johann Wolfgang University, Frankfurt/Main, Germany
- Institute of Phonetics, Johann Wolfgang University, Frankfurt/Main, Germany
- Institute of Acoustics and Speech Communication, Dresden University of Technology, Dresden, Germany
- International Institute for Advanced Scientific Studies, Italy
- Second University of Naples, Caserta, Italy
- Regione Campania, Italy
- Provincia di Salerno, Italy

Table of Contents

I Emotions and ICT

Cross-Fertilization between Studies on ICT Practices of Use and Cross-Modal Analysis of Verbal and Nonverbal Communication 1
Leopoldina Fortunati, Anna Esposito, and Jane Vincent

Theories without Heart 5
Leopoldina Fortunati

Prosodic Characteristics and Emotional Meanings of Slovak Hot-Spot Words 18
Štefan Beňuš and Milan Rusko

Affiliations, Emotion and the Mobile Phone 28
Jane Vincent

Polish Emotional Speech Database – Recording and Preliminary Validation 42
Piotr Staroniewicz and Wojciech Majewski

Towards a Framework of Critical Multimodal Analysis: Emotion in a Film Trailer 50
Maria Bortoluzzi

Biosignal Based Emotion Analysis of Human-Agent Interactions 63
Evgenia Hristova, Maurice Grinberg, and Emilian Lalev

Emotional Aspects in User Experience with Interactive Digital Television: A Case Study on Dyslexia Rehabilitation 76
Filomena Papa and Bartolomeo Sapio

Investigation of Normalised Time of Increasing Vocal Fold Contact as a Discriminator of Emotional Voice Type 90
Peter J. Murphy and Anne-Maria Laukkanen

Evaluation of Speech Emotion Classification Based on GMM and Data Fusion 98
Martin Vondra and Robert Vích

Spectral Flatness Analysis for Emotional Speech Synthesis and Transformation 106
Jiří Přibil and Anna Přibilová

II Verbal and Nonverbal Features of Computational Phonetics

Voice Pleasantness of Female Voices and the Assessment of Physical Characteristics 116
Vivien Zuta

Technical and Phonetic Aspects of Speech Quality Assessment: The Case of Prosody Synthesis 126
Jana Tučková, Jan Holub, and Tomáš Duběda

Syntactic Doubling: Some Data on Tuscan Italian 133
Anna Esposito

Perception of Czech in Noise: Stability of Vowels 149
Jitka Veroňková and Zdena Palková

Challenges in Segmenting the Czech Lateral Liquid 162
Radek Skarnitzl

Implications of Acoustic Variation for the Segmentation of the Czech Trill /r/ 173
Pavel Machač

Voicing in Labial Plosives in Czech 182
Annett B. Jorschick

Normalization of the Vocalic Space 190
Jan Volín

III Algorithmic and Theoretical Analysis of Multimodal Interfaces

Gaze Behaviors for Virtual Crowd Characters 201
Helena Grillon, Barbara Yersin, Jonathan Maïm, and Daniel Thalmann

Gestural Abstraction and Restatement: From Iconicity to Metaphor 214
Nicla Rossini

Preliminary Prosodic and Gestural Characteristics of Instructing Acts in Polish Task-Oriented Dialogues 227
Maciej Karpiński

Polish Children's Gesticulation in Narrating (Re-telling) a Cartoon 239
Ewa Jarmołowicz-Nowikow

Prediction of Learning Abilities Based on a Cross-Modal Evaluation of Non-verbal Mental Attributes Using Video-Game-Like Interfaces 248
Yiannis Laouris, Elena Aristodemou, and Pantelis Makris

Automatic Sentence Modality Recognition in Children's Speech, and Its Usage Potential in the Speech Therapy 266
Dávid Sztahó, Katalin Nagy, and Klára Vicsi

Supporting Engagement and Floor Control in Hybrid Meetings 276
Rieks op den Akker, Dennis Hofs, Hendri Hondorp, Harm op den Akker, Job Zwiers, and Anton Nijholt

Behavioral Consistency Extraction for Face Verification 291
Hui Fang and Nicholas Costen

Protecting Face Biometric DCT Templates by Means of Pseudo-random Permutations 306
Marcos Faundez-Zanuy

Facial Expressions Recognition from Image Sequences 315
Zahid Riaz, Christoph Mayer, Michael Beetz, and Bernd Radig

Czech Artificial Computerized Talking Head George 324
Josef Chaloupka and Zdenek Chaloupka

An Investigation into Audiovisual Speech Correlation in Reverberant Noisy Environments 331
Simone Cifani, Andrew Abel, Amir Hussain, Stefano Squartini, and Francesco Piazza

Articulatory Speech Re-synthesis: Profiting from Natural Acoustic Speech Data 344
Dominik Bauer, Jim Kannampuzha, and Bernd J. Kröger

A Blind Source Separation Based Approach for Speech Enhancement in Noisy and Reverberant Environment 356
Alessio Pignotti, Daniele Marcozzi, Simone Cifani, Stefano Squartini, and Francesco Piazza

Quantitative Analysis of the Relative Local Speech Rate 368
Jan Janda

Czech Spontaneous Speech Collection and Annotation: The Database of Technical Lectures 377
Josef Rajnoha and Petr Pollák

BSSGUI – A Package for Interactive Control of Blind Source Separation Algorithms in MATLAB 386
Jakub Petkov and Zbyněk Koldovský

Accuracy Analysis of Generalized Pronunciation Variant Selection in ASR Systems 399
Václav Hanžl and Petr Pollák

Analysis of the Possibilities to Adapt the Foreign Language Speech Recognition Engines for the Lithuanian Spoken Commands Recognition 409
Rytis Maskeliunas, Algimantas Rudzionis, and Vytautas Rudzionis

MLLR Transforms Based Speaker Recognition in Broadcast Streams . . . 423
Jan Silovsky, Petr Cerva, and Jindrich Zdansky

Author Index 433