

Lecture Notes in Artificial Intelligence 5078

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

Subseries of Lecture Notes in Computer Science

Elisabeth André Laila Dybkjær
Wolfgang Minker Heiko Neumann
Roberto Pieraccini Michael Weber (Eds.)

Perception in Multimodal Dialogue Systems

4th IEEE Tutorial and Research Workshop on Perception and
Interactive Technologies for Speech-Based Systems, PIT 2008
Kloster Irsee, Germany, June 16-18, 2008
Proceedings

Volume Editors

Elisabeth André

Universität Augsburg, Institut für Informatik, Augsburg, Germany

E-mail: andre@informatik.uni-augsburg.de

Laila Dybkjær

Prolog Development Center A/S (PDC), Brøndby, Denmark

E-mail: laila@pdc.dk

Wolfgang Minker

Universität Ulm, Institut für Informationstechnologie und Dialogsysteme

Ulm, Germany

E-mail: wolfgang.minker@uni-ulm.de

Heiko Neumann

Universität Ulm, Abteilung Neuroinformatik, Ulm, Germany

E-mail: heiko.neumann@uni-ulm.de

Roberto Pieraccini

SpeechCycle, Inc., New York, NY, USA

E-mail: roberto@speechcycle.com

Michael Weber

Universität Ulm, Institut für Medieninformatik, Ulm, Germany

E-mail: michael.weber@uni-ulm.de

Library of Congress Control Number: 2008928841

CR Subject Classification (1998): H.5.2-3, H.5, I.2.6-7, I.2.10, I.2, K.4, I.4

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-540-69368-8 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-69368-0 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 is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2008

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper SPIN: 12322121 06/3180 5 4 3 2 1 0

Preface

The IEEE Tutorial and Research Workshop on Perception and Interactive Technologies for Multimodal Dialogue Systems (PIT 2008) is the continuation of a successful series of workshops that started with an ISCA Tutorial and Research Workshop on Multimodal Dialogue Systems in 1999. This workshop was followed by a second one focusing on mobile dialogue systems (IDS 2002), a third one exploring the role of affect in dialogue (ADS 2004), and a fourth one focusing on perceptive interfaces (PIT 2006). Like its predecessors, PIT 2008 took place at Kloster Irsee in Bavaria.

Due to the increasing interest in perceptive interfaces, we decided to hold a follow-up workshop on the themes discussed at PIT 2006, but encouraged above all papers with a focus on perception in multimodal dialogue systems. PIT 2008 received 37 papers covering the following topics (1) multimodal and spoken dialogue systems, (2) classification of dialogue acts and sound, (3) recognition of eye gaze, head poses, mimics and speech as well as combinations of modalities, (4) vocal emotion recognition, (5) human-like and social dialogue systems and (6) evaluation methods for multimodal dialogue systems.

Noteworthy was the strong participation from industry at PIT 2008. Indeed, 17 of the accepted 37 papers come from industrial organizations or were written in collaboration with them.

We would like to thank all authors for the effort they made with their submissions, and the Program Committee – nearly 50 distinguished researchers from industry and academia – who worked very hard to meet tight deadlines and selected the best contributions for the final program. Special thanks goes to our invited speaker, Anton Batliner from Friedrich-Alexander-Universität Erlangen-Nürnberg.

A number of organizations supported PIT 2008 including the IEEE Signal Processing Society, ACL Sigmedia and ACL/ISCA Sigdial. In particular, we are grateful to Springer for publishing the proceedings in their LNCS/LNAI series.

April 2008

Elisabeth André
Laila Dybkjær
Wolfgang Minker
Heiko Neumann
Roberto Pieraccini
Michael Weber

Organization

Organizing Committee

Elisabeth André (University of Augsburg, Germany)
Laila Dybkjær (PDC A/S, Denmark)
Wolfgang Minker (University of Ulm, Germany)
Roberto Pieraccini (SpeechCycle Inc., USA)
Heiko Neumann (University of Ulm, Germany)
Michael Weber (University of Ulm, Germany)

Scientific Committee

Jan Alexandersson	Paul Heisterkamp	Christopher Peters
Yacine Bellik	Dirk Heylen	Helmut Prendinger
André Berton	David House	Norbert Reithinger
Rolf Carlson	Christophe Jacquet	Thomas Rist
Cristina Conati	Ralf Kompe	Laurent Romary
Klaus Dorfmueller-Ulhaas	Rainer Lienhart	Fiorella de Rosis
Ellen Douglas-Cowie	Jackson Liscombe	Alex Rudnicky
Gabriel Cristobal	Joseph Mariani	Candy Sidner
Laurence Devillers	Jean-Claude Martin	Rainer Stiefelhagen
Sadaoki Furui	Dominic Massaro	Oliviero Stock
Jim Glass	Louis-Philippe Morency	David Suendermann
Silke Goronzy	Elmar Nöth	David Traum
Joakim Gustafson	Ana Paiva	Matthias Thomae
Eli Hagen	Tim Paek	Thurid Vogt
Hani Hagrais	Patrick Paroubek	Wayne Ward
Udo Haiber	Catherine Pelachaud	
Gerhard Hanrieder	Alex Pentland	

Sponsoring Organizations

IEEE Signal Processing Society
ACL/ISCA Sigdial
ACL Sigmedia

Table of Contents

Invited Keynote

Whence and Whither: The Automatic Recognition of Emotions in Speech	1
<i>Anton Batliner</i>	

Multimodal and Spoken Dialogue Systems

A Generic Spoken Dialogue Manager Applied to an Interactive 2D Game	2
<i>Andrea Corradini and Christer Samuelsson</i>	
Adaptive Dialogue Management in the NIMITEK Prototype System ...	14
<i>Milan Gnjatović and Dietmar Rösner</i>	
Adaptive Search Results Personalized by a Fuzzy Recommendation Approach	26
<i>Philipp Fischer, André Berton, and Andreas Nürnberger</i>	
Factors Influencing Modality Choice in Multimodal Applications	37
<i>Anja B. Naumann, Ina Wechsung, and Sebastian Möller</i>	
Codebook Design for Speech Guided Car Infotainment Systems	44
<i>Martin Raab, Rainer Gruhn, and Elmar Noeth</i>	
Evaluating Text Normalization for Speech-Based Media Selection	52
<i>Martin Pfeil, Dirk Buehler, Rainer Gruhn, and Wolfgang Minker</i>	

Classification of Spoken Utterances and Sound

A Two Phases Statistical Approach for Dialog Management	60
<i>David Griol, Lluís F. Hurtado, Encarna Segarra, and Emilio Sanchis</i>	
Detecting Problematic Dialogs with Automated Agents	72
<i>Alexander Schmitt, Carolin Hank, and Jackson Liscombe</i>	
Call Classification with Hundreds of Classes and Hundred Thousands of Training Utterances..... and No Target Domain Data	81
<i>David Suendermann, Phillip Hunter, and Roberto Pieraccini</i>	
Hard vs. Fuzzy Clustering for Speech Utterance Categorization	88
<i>Amparo Albalade and David Suendermann</i>	

Static and Dynamic Modelling for the Recognition of Non-verbal Vocalisations in Conversational Speech	99
<i>Björn Schuller, Florian Eyben, and Gerhard Rigoll</i>	

Recognition of Eye Gaze, Head Pose, Mimics and Lip Movements

Writing with Your Eye: A Dwell Time Free Writing System Adapted to the Nature of Human Eye Gaze	111
<i>Nikolaus Bee and Elisabeth André</i>	
Unsupervised Learning of Head Pose through Spike-Timing Dependent Plasticity	123
<i>Ulrich Weidenbacher and Heiko Neumann</i>	
Spoken Word Recognition from Side of Face Using Infrared Lip Movement Sensor	132
<i>Takahiro Yoshida, Erika Yamazaki, and Seiichiro Hangai</i>	
Neurobiologically Inspired, Multimodal Intention Recognition for Technical Communication Systems (NIMITEK)	141
<i>Andreas Wendemuth, Jochen Braun, Bernd Michaelis, Frank Ohl, Dietmar Rösner, Henning Scheich, and Ralf Warnemünde</i>	

Speech Recognition

Deploying DSR Technology on Today's Mobile Phones: A Feasibility Study	145
<i>Dmitry Zaykovskiy and Alexander Schmitt</i>	
Real-Time Recognition of Isolated Vowels	156
<i>Mara Carvalho and Aníbal Ferreira</i>	
Improving Robustness in Jacobian Adaptation for Noisy Speech Recognition	168
<i>Yongjoo Jung</i>	
Comparing Linear Feature Space Transformations for Correlated Features	176
<i>Daniel Vásquez, Rainer Gruhn, Raymond Brueckner, and Wolfgang Minker</i>	

Vocal Emotion Recognition and Annotation

EmoVoice—A Framework for Online Recognition of Emotions from Voice	188
<i>Thurid Vogt, Elisabeth André, and Nikolaus Bee</i>	

Real-Time Emotion Recognition Using Echo State Networks	200
<i>Stefan Scherer, Mohamed Oubbati, Friedhelm Schwenker, and Günther Palm</i>	
Emotion Classification of Audio Signals Using Ensemble of Support Vector Machines	205
<i>Taner Danisman and Adil Alpkocak</i>	
On the Influence of Phonetic Content Variation for Acoustic Emotion Recognition	217
<i>Bogdan Vlasenko, Björn Schuller, Andreas Wendemuth, and Gerhard Rigoll</i>	
On the Use of Kappa Coefficients to Measure the Reliability of the Annotation of Non-acted Emotions	221
<i>Zoraida Callejas and Ramón López-Cózar</i>	
Annotation of Emotion in Dialogue: The Emotion in Cooperation Project	233
<i>Federica Cavicchio and Massimo Poesio</i>	

Human-Like Social Dialogue

Potential Benefits of Human-Like Dialogue Behaviour in the Call Routing Domain	240
<i>Joakim Gustafson, Mattias Heldner, and Jens Edlund</i>	
Human-Likeness in Utterance Generation: Effects of Variability	252
<i>Anna Hjalmarsson and Jens Edlund</i>	
Designing Socially Aware Conversational Agents	256
<i>Koen van Turnhout, Jacques Terken, and Berry Eggen</i>	
A Prototype for Future Spoken Dialog Systems Using an Embodied Conversational Agent	268
<i>Marcel Dausend and Ute Ehrlich</i>	
Innovative Interfaces in MonAMI: The Reminder	272
<i>Jonas Beskow, Jens Edlund, Björn Granström, Joakim Gustafson, and Gabriel Skantze</i>	

Evaluation Methods

Evaluation Methods for Multimodal Systems: A Comparison of Standardized Usability Questionnaires	276
<i>Ina Wechsung and Anja B. Naumann</i>	

Subjective Evaluation Method for Speech-Based Uni- and Multimodal Applications.....	285
<i>Benjamin Weiss, Ina Wechsung, Anja Naumann, and Sebastian Möller</i>	
Weighting the Coefficients in PARADISE Models to Increase Their Generalizability	289
<i>Klaus-Peter Engelbrecht, Christine Kühnel, and Sebastian Möller</i>	
EXPROS: A Toolkit for Exploratory Experimentation with Prosody in Customized Diphone Voices	293
<i>Joakim Gustafson and Jens Edlund</i>	
Automatic Evaluation Tool for Multimodal Dialogue Systems.....	297
<i>Hugo Wesseling, Matthias Bezold, and Nicole Beringer</i>	
Towards a Perception-Based Evaluation Model for Spoken Dialogue Systems	306
<i>Sebastian Möller and Klaus-Peter Engelbrecht</i>	
Author Index	311