

Newsletter

www.amiproject.org

AMI c/o Idiap Research Institute, Centre du Parc,
Rue Marconi 19, P.O. Box 592, CH-1920 Martigny
info@amiproject.org - www.amiproject.org

Contents

COVER STORY

- Noldus information technology joins AMIDA 1

FOCUS

- Completed PhD Thesis, University of Edinburgh 2
- Keyword spotting in security and defense applications 3

INSIDE AMI

- Events 4
- Selected publications 4

Events

ICMI-MLMI 2009

Nov 2-6, 2009
Cambridge, USA

The 11th International Conference on Multimodal Interfaces and Workshop on Machine Learning for Multi-modal Interaction

<http://icmi2009.acm.org>

ASRU 2009

December 13-17, 2009
Merano, Italy

The 11th biannual IEEE workshop on Automatic Speech Recognition and Understanding

<http://www.asru2009.org>

© AMI Consortium, all rights reserved

Noldus information technology joins AMIDA

NEW TECHNOLOGY PARTNER IN AMIDA PROJECT

Noldus Information Technology (www.noldus.com) develops and markets professional software and instrumentation for research on human or animal behavior. Our products enable the collection, analysis, management, and presentation of behavioral data.

Scientists, engineers, and practitioners use our products to study behavioral processes, automate measurements, improve the quality of their data, increase productivity, and make optimal use of human or animal resources. Our systems have found their way to more than 4.500 universities, research institutes, and companies in over 75 countries. Applications are found in a wide range of disciplines, from neuroscience to human-computer interaction, and from psychology to sports research.

Our product range includes a number of products developed for human behavior research, such as The Observer XT, Pocket Observer, Theme, FaceReader, and uLog. In addition, we offer accessories to complete an experimental set-up, such as specialized data-entry keyboards, eye trackers, mobile device cameras, and spectacles cameras. Next to that, we can provide integrated data acquisition and data analysis systems, including PCs and various sorts of audiovisual equipment, as well as complete fixed or portable observation labs.

Besides the sales of products, we offer consulting, demonstration, and rental services, technical support, and training. Next to our headquarters in The Netherlands, we operate from several regional offices in the USA, China and Europe, as well as through

Noldus
Information Technology

a worldwide network of distributors. Our team of 100 enthusiastic professionals are ready to help you.

Noldus will use The Observer XT 9.0 as the foundation for the work to be carried out in the framework of AMIDA. The Observer (www.noldus.com/observer) is Noldus' flagship software package for collection and analysis of behavioral data. The Observer has been in continuous development since the start of the company in 1989. Originally designed as a tool for recording and analyzing observed events with a computer, the package evolved into a tool for the annotation of a wide variety of behavioral data from digital video (and audio) streams

We will extend The Observer XT with specific functions that automate part of the annotation process, like speaker identification, speech recognition, disfluency correction and gesture recognition. Those functions will be based on algorithms and software components developed by AMIDA partners. For this purpose, Noldus will develop a generic plug-in mechanism through which third-party software components (e.g. an automatic speech recognition module or a gesture recognizer) can be added to the package, to process audio and video signals in real-time or offline.

To be continued on page 2

Cover Story

Noldus information technology joins AMIDA (continued)

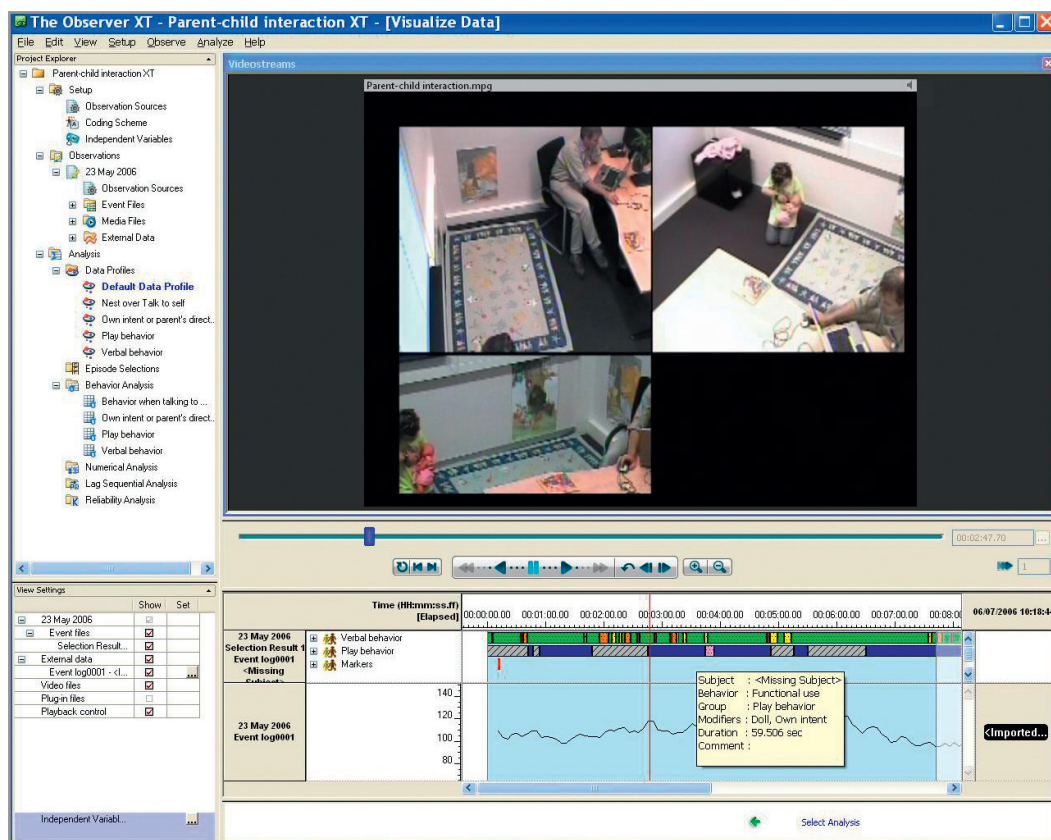
This will then give the user the best of both worlds: cutting-edge automated behavior recognition by AMIDA technology and state-of-the-art annotation for all other behaviors (for which no automated recording exists yet) with The Observer software.

There are several use cases where The Observer XT with the AMIDA framework can be used and which represent applications with substantial societal and economic relevance, for example: eating and drinking behavior, usability testing, user experience studies and focus groups.

Contact:

Noldus Information Technology
P.O. Box 268
6700 AG Wageningen
The Netherlands
Phone +31-317-473300
Fax +31-317-424496
E-mail info@noldus.nl
Web www.noldus.com

Erik Meyer
E.Meyer@Noldus.NL



Completed PhD Thesis, Alfred Dielmann

AUTOMATIC RECOGNITION OF MULTIPARTY HUMAN INTERACTIONS USING DYNAMIC BAYESIAN NETWORKS

Alfred Dielmann started his PhD at the University of Edinburgh under the supervision of Prof. Steve Renals in January 2004, and successfully defended his thesis in October 2008.

His PhD research, conducted in the framework of the AMI and AMIDA EU projects, investigated automatic meeting segmentation both in terms of Dialogue Acts (DAs) and Meeting Actions. DAs model the discourse structure at a fine grained level, highlighting individual speaker intentions. Group meeting actions describe the same process at a coarse level, highlighting interactions between different meeting participants and showing overall group intentions.

A framework based on dynamic Bayesian networks (DBNs) as investigated for both tasks. A first set of experiments is concerned with the segmentation and structuring of meetings (recorded using multiple cameras and microphones) into sequences of group meeting

actions such as monologue, discussion and presentation. Four families of multimodal features were related to complex group behaviours using a multi-stream modelling framework based on DBNs. Later experiments are concerned with the automatic recognition of DAs in multiparty conversational speech, proposing a switching DBN approach in which segmentation and classification of DAs are carried out in parallel.

The DBN based approach, when applied both to the meeting action and the dialogue act recognition task, yielded significant improvements over baseline Hidden Markov Models. On both tasks, the DBN framework provided an effective factorisation of the state-space and a flexible infrastructure able to integrate a heterogeneous set of resources such as continuous and discrete multimodal features, and statistical language models.



Alfred Dielmann

Alfred Dielmann
Idiap Research Institute
alfred.dielmann@idiap.ch

Keyword spotting developed with the help of AMIDA finds its way to security and defense applications

THE SPEECH TEAM OF BRNO UNIVERSITY OF TECHNOLOGY IN THE SECURITY AND DEFENSE DOMAINS

One of the key tasks of Brno University of Technology (BUT) speech team in AMIDA is to research and develop techniques for keyword spotting and spoken term detection in spontaneous speech. In parallel with work on meeting data, BUT group has investigated analysis of conversational telephone speech (CTS) for security and defense applications. CTS was also one task in NIST Spoken term detection evaluations in 2006, in which BUT successfully participated.

Since 2007, BUT cooperates with two other Czech speech research groups: at Technical University of Liberec (TUL), and at University of West Bohemia (UWB), on a research project "Overcoming the language barrier complicating investigation into financing terrorism and serious financial crimes" (Czech Ministry of Interior project No. VD20072010B16). The project aims at the analysis of spontaneous telephone calls from security and defense domains. Unlike English, where corpora such as Switchboard and Fisher provide sufficient amounts of training material, Czech lacked a well transcribed database of spontaneous telephone calls. This is why in 2007, the activities of the project concentrated on the creation of such a database – the consortium now holds almost 100 hours of transcribed and checked spontaneous speech data.

To compare the performances of systems, an evaluation of keyword spotting systems was organized in 2008, with the final run in November. Systems were compared using standard metrics such as FOM (figure of merit) and EER (equal error

rate), and their speed and ability to handle OOV (out of vocabulary) words were also compared.

BUT tested 4 systems in this evaluation: FastLVCSR was based on LVCSR with insertion of keywords into language model; HybridLVCSR used full-fledged word and subword recognition and indexing; and two acoustic systems were based on GMM/HMM and NN/HMM. While LVCSR systems are more precise, the advantage of acoustic ones is in their speed. HybridLVCSR is worth mentioning as it allows for pre-processing large quantities of data off-line with subsequent very fast searches, including OOVs.

In 2009, units of Czech Ministries of Defense and Interior are extensively testing the keyword spotting engines and provide the researchers with very useful feedback concerning robustness to environmental noise, score calibration, user interfaces and applicability in real security scenarios.

For more information about this work, please contact Honza Cernocky cernocky@fit.vutbr.cz, Igor Szoke szoke@fit.vutbr.cz or Michal Fapso ifapso@fit.vutbr.cz. Further information is published on the website of Speech Processing Group at Faculty of Information Technology, Brno University of Technology: <http://speech.fit.vutbr.cz/en/media/first-evaluation-keyword-spotting-czech-supported-ministries-interior-and-defense>

Honza Cernocky
Brno University of Technology



Visualization of BUT keyword spotting results

Events

Joint ICMI-MLMI 2009

Nov 2-6, 2009, Cambridge, USA

The 11th International Conference on Multimodal Interfaces (ICMI) and the 6th International Workshop on Machine Learning for Multi-modal Interaction (MLMI) will jointly take place at the MIT Media Lab on November 2-6, 2009.

ICMI-MLMI aims to further scientific research in multimodal interaction, methods and systems.

The joint conference will feature a single-track main conference with keynote speakers, panel discussions, technical paper presentations, poster sessions, demonstrations, and workshops.

Organizing Committee

General Co-chairs:

- James Crowley, INRIA
- Yuri Ivanov, MERL
- Christopher Wren, Google

Program Co-Chairs:

- Daniel Gatica-Perez, Idiap Research Institute
- Michael Johnston, AT&T Research
- Rainer Stiefelhagen, University of Karlsruhe

Important Dates:

Author notification: July 20, 2009

Conference: November 2-4, 2009

More information: <http://icmi2009.acm.org>

Daniel Gatica-Perez
Idiap Research Institute
gatica@idiap.ch

IEEE ASRU2009

December 13-17, 2009, Merano, Italy

The 11th biannual IEEE workshop on Automatic Speech Recognition and Understanding (ASRU) will be held on December 13-17, 2009. The ASRU workshops have a tradition of bringing together researchers from academia and industry in an intimate and collegial setting to discuss problems of common interest in automatic speech recognition and understanding.

Workshop topics:

- automatic & human speech recognition and understanding
- speech to text systems
- spoken dialog systems
- multilingual language processing

- robustness in ASR
- spoken document retrieval
- speech-to-speech translation
- spontaneous speech processing
- speech summarization
- new applications of ASR

Important dates:

15.07.09: Paper submission deadline
03.09.09: Paper notification of acceptance
24.09.09: Demo session proposal deadline
07.10.09: Early registration deadline
13-17.12.2009: Workshop

General Chairs

Giuseppe Riccardi, U. Trento, Italy
Renato De Mori, U. Avignon, France

Technical Chairs

Jeff Bilmes, U. Washington, USA
Pascale Fung, HKUST, Hong Kong China
Shri Narayanan, USC, USA
Tanja Schultz, U. Karlsruhe, Germany

Registration will be handled via the ASRU 2009 website, <http://www.asru2009.org>, where more information on the workshop will be available.

Selected publications

Comparing word, character, and phoneme n-grams for subjective utterance recognition.

T. Wilson and S. Raaijmakers
In *Interspeech 2008*, Brisbane, Australia, 2008.

Decision-Level Fusion for Audio-Visual Laughter Detection.

B. Reuderink, M. Poel, K. P. Truong, R. Poppe and M. Pantic
In *5th Joint Workshop on Machine Learning and Multimodal Interaction, MLMI 2008*, pages 137-148, Springer Verlag, 2008.

Design and Evaluation of Systems to Support Interaction Capture and Retrieval.

S. Whittaker, S. Tucker, K. Swampillai, R. Laban
In *Personal and Ubiquitous Computing*, volume 12, number 3, pages 197-221, 2008.

Designing Awareness Support for Distributed Cooperative Design Teams.

D. Vyas, D. Heylen and A. Nijholt
In *15th European Conference on Cognitive Ergonomics*, pages 23-26, ACM, 2008.

Determining Latency for on-line Dialog Act Classification.

S. Gemesin
In *5th Joint Workshop on Machine Learning and Multimodal Interaction, MLMI'08*, 2008.

Discriminative human action recognition using pairwise CSP classifiers.

R. Poppe and M. Poel
In *8th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008)*, 2008.

Domain-specific Classification Methods for Disfluency Detection.

S. Gemesin, T. Becker and P. Poller
In *Interspeech 2008*, Brisbane, Australia.

Effect of sound spatialisation on multitasking in remote meetings.

S. N. Wrigley, S. Tucker, G. J. Brown and S. Whittaker
In *Proceedings of Acoustics'08*, 2008, Paris

Estimating the Dominant Person in Multi-Party Conversations Using Speaker Diarization Strategies.

H. Hung, Y. Huang, G. Friedland and D. Gatica-Perez
In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.

Exploiting 'Subjective' Annotations.

D. Reidsma and H. J. A. op den Akker
In *Coling 2008: Proceedings of the workshop on Human Judgements in Computational Linguistics*, pages 8-16, Coling 2008 Organizing Committee, 2008.

Exploring Features and Classifiers for Dialogue Act Segmentation.

H. op den Akker and C. Schulz
In *Machine Learning for Multimodal Interaction, MLMI 2008*, Utrecht, the Netherlands, pages 196-207, Springer Verlag, Utrecht, the Netherlands, 2008.

Exploring Mediated Interactions: A Design Exercise.

D. Vyas, Y. Liu and A. Nijholt
In *15th European Conference on Cognitive Ergonomics*, pages 1-2, ACM, 2008.

Extrinsic Summarization Evaluation: A Decision Audit Task.

G. Murray, T. Kleinbauer, P. Poller, S. Renals, J. Kilgour and T. Becker
In *Machine Learning for Multimodal Interaction - 5th International Workshop, MLMI 2008*, pages 349-361, 2008.

Filter Bank Design Based on Minimization of Individual Aliasing Terms for Minimum Mutual Information Subband Adaptive Beamforming.

K. Kumatani, J. McDonough, S. Schacht, D. Klakow, P. N. Garner and W. Li
In *Proceedings International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, pages 1609-1612, 2008.

How do I address you? Modelling addressing behaviour based on an analysis of multi-modal corpora of conversational discourse.

R. op den Akker and M. Theune
In *AISB 2008 Symposium on Multimodal Output Generation (MOG 2008)*, pages 10-17, Aberdeen, UK, 2008.

Hybrid Multi-Step Disfluency Detection.

S. Gemesin, T. Becker and P. Poller
5th Joint Workshop on Machine Learning and Multimodal Interaction