

# AMI Newsletter

## Contents

### COVER STORY

AMI's Participation in NIST's RT-05S Evaluation 1

### FOCUS

• DFKI 2  
• AMI's Participation in NIST's RT-05S Evaluation 3

### INSIDE AMI

• ICMi 2005 4  
• Journal of Machine Learning Research 4  
• Workshop Harem 2005 4

## News

2005-07-05

MLMI'05 will be held on 11-13 July 2005, Royal College of Physicians, Edinburgh, UK

2005-07-05

The 6<sup>th</sup> issue of the AMI Newsletter is online.

© AMI Integrated Project, all rights reserved

## AMI's Participation in NIST's RT-05S Evaluation

In May 2005, NIST held its Rich Transcription Spring'05 evaluation, RT-05S, which targets the automatic processing of Meetings data. This was the third in a series of (roughly) annual evaluations that NIST has held to benchmark core recognition capabilities in speech processing for the Meetings domain. This year's evaluation represented a significant expansion over past years, this time featuring four different tasks and two different Meetings data types. The evaluation has traditionally focused on «conference room» Meeting data, of the sort AMI is collecting, but this year also introduced a «lecture room» track, and the task list expanded from speech transcription (STT, for speech-to-text) and speaker segmentation & tracking (SPKR) to also include speech activity detection (SAD) and source localization (SLOC).

AMI was a major contributor to this year's evaluation, both as a system builder and as a data provider. AMI labs participated in 3 of the 4 tasks (all but SLOC, which was not offered in the conference room track) and fielded a number of high-performance systems. In addition, the consortium contributed a sampling from its on-going data collection to provide test material for the evaluation, enriching the mix of meeting rooms and recording conditions evaluated and allowing participating sites to explore this new data collection.

### The AMI Speech Transcription System

This was a first opportunity to showcase the Meetings automatic speech recognition (ASR) system being developed expressly for the AMI project, under the direction of Thomas Hain of University of Sheffield. This massive development effort, launched in spring '04, was a true consortium-level project, including participation from a half dozen different labs jointly building an ASR system from scratch. The system uses the HTK

toolkit (and extensions) developed by Cambridge University and takes advantage of a wide range of speech and language resources, including conversational speech collections such as the Switchboard corpus of recorded telephone calls and Meetings data collected by ICSI, CMU, and NIST, and now incorporating the data being recorded by AMI.

This ambitious and widely distributed system building effort was organized into several packages. The work groups (with host site noted in parentheses) focused on:

- dictionary development (IDIAP) - assembling the recognizer's wordlists and generating pronunciations
- language modelling (Sheffield) - training a variety of models encoding word sequencing information
- audio pre-processing (IDIAP) - preparing the raw audio stream by performing speech/silence segmentation and dealing with different types of microphones
- core (Sheffield) - generating basic acoustic models and managing the recognizer decoding process
- adaptation (Edinburgh) - modifying acoustic models from speaker-independent baseforms to speaker-sensitive ones

Most of these groups themselves included participation across multiple sites; in addition to the labs noted above, Brno, Twente, and ICSI all made substantial contributions to this effort. The team was able to make rapid progress, attaining highly competitive performance in less than a year from project start. The system currently achieves word error rates of 32.7% on last year's RT-04S Meetings evaluation set (which includes meetings recorded at NIST, CMU, LDC, and ICSI) when testing on close-talking high-quality headsets and 44.7% using the far-field tabletop microphones.

To be continued on page 3

Cover Story

## DFKI

GERMAN RESEARCH CENTER FOR ARTIFICIAL INTELLIGENCE GMBH



Founded in 1988, DFKI today is one of the largest nonprofit contract research institutes in the field of innovative software technology based on Artificial Intelligence (AI) methods. DFKI is focusing on the complete cycle of innovation - from world-class basic research and technology development through leading-edge demonstrators and prototypes to product functions and commercialization. Based in Kaiserslautern and Saarbrücken, the German Research Center for Artificial Intelligence ranks among the «Centers of Excellence» worldwide.

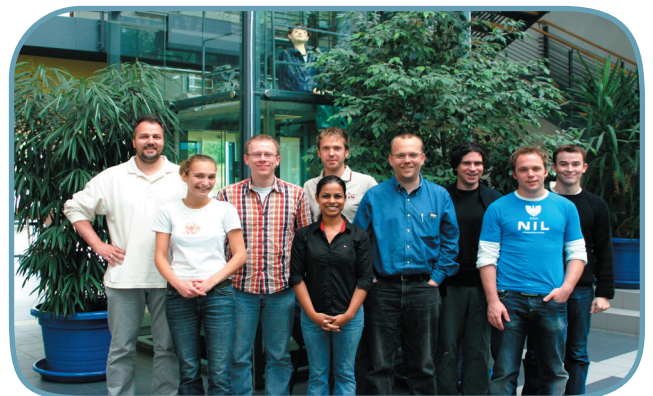
DFKI's six research departments are directed by internationally recognized research scientists: Image Understanding and Pattern Recognition (Director: Prof. Thomas Breuel), Knowledge Management (Director: Prof. Andreas Dengel), Intelligent Visualization and Simulation Systems (Director: Prof. Hans Hagen), Deduction and Multiagent Systems (Director: Prof. Jörg Siekmann), Language Technology (Director: Prof. Hans Uszkoreit), and Intelligent User Interfaces (Director and CEO of DFKI: Prof. Wolfgang Wahlster). Furthermore, since 2002 the Institute for Information Systems (IWi) is part of the DFKI.

The mission of four Transfer Centers is to transfer DFKI's research results to commercial application and there are currently three Competence Centers, that bundle know-how, skills, and technologies of DFKI.

DFKI benefits from interaction with the faculty of the Universities of Saarbrücken and Kaiserslautern and in turn provides opportunities for research and Ph.D. thesis supervision to students from these universities, which have an outstanding reputation in Computer Science.

## DFKI in AMI

Work on the AMI project at DFKI comes from a group of researchers, software engineers and students in the Intelligent User Interfaces department (see the group picture). Their research focuses on the multimodal structure and content analysis of meetings. The expertise covers multiple areas: Natural Language Understanding, Discourse Modeling, and Natural Language Generation are the key for the tasks in AMI's workpackage 5 including segmentation, information structuring, and summarization of meetings.



*The AMI team (from left): Tilman Becker (project leader at DFKI), Kateryna Mysak, Harald Lochert, Sujatha Balachandran, Thomas Kleinbauer, Jan Alexandersson, Christoph Lauer, Jochen Frey, and Benjamin Lang.*

The technologies used include statistical methods (e.g., for the segmentation and classification of dialogue acts) as well as symbolic approaches (e.g., the development of an ontology for the design meetings of the AMI hub-scenario). Another contribution is the development of AMIGram, an annotation and browsing tool for the corpora collected and annotated in the AMI project.

DFKI is also managing AMI's work on technology transfer. On the web you can find more information about AMI's first technology transfer event that took place in Brussels this spring:  
<http://www.amiproject.org/TTW05/index.php>

On the web: <http://www.dfki.de> - contact: [Tilman.Becker@dfki.de](mailto:Tilman.Becker@dfki.de)



*Prof. Dr. Wolfgang Wahlster, CEO of DFKI and head of the Intelligent User Interfaces department.*

## AMI's Participation in NIST's RT-05S Evaluation

CONTINUED FROM PAGE 1

Word error rates on development sets from the newly-collected AMI data range between 30-35% for close-talking mics and 35-40% with tabletop microphones.

More information on the AMI ASR system will appear in «The Development of the AMI System for the Transcription of Speech in Meetings» at MLMI'05 in July, and «Transcription of Conference Room Meetings: An Investigation» at Interspeech'05 in September.

### Other AMI Participation

In addition to the AMI-ASR multi-site team submission, several AMI labs also submitted evaluation systems of their own.

ICSI, in collaboration with long-time research partner SRI International, continued its involvement in the RT-S evaluation series. This is ICSI's third year of participation on the STT task, and their first submission to the SPKR task for the Meetings domain. The speech transcription system builds on SRI's DECIPHER speech recognition system, incorporating joint ICSI/SRI developments in the telephone and broadcast news domains and now introducing processing specifically targeting challenges in Meetings

recognition. The new speaker segmentation and tracking system, based on algorithms originally developed by IDIAP, incorporates such new features as delay-sum processing for effectively integrating data from multiple tabletop microphones.

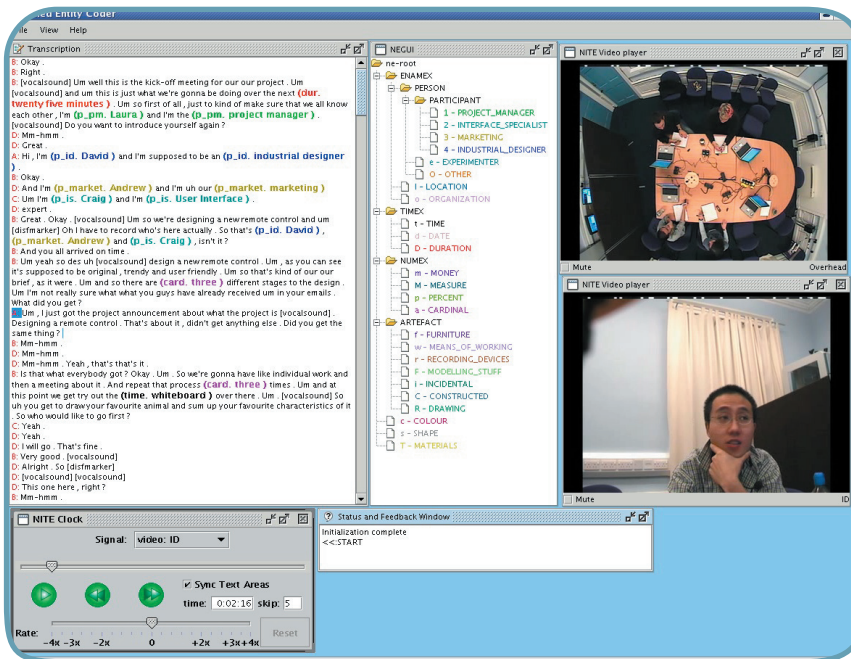
TNO, which has been a regular participant in NIST's Speaker Recognition evaluations, this year participated in the Rich Transcription evaluation series for the first time. The TNO team, led by David van Leeuwen, submitted systems in both the SPKR and SAD tasks, building a two-model speech vs. silence decoder for the SAD task and further processing the speech portions using speaker segmentation and clustering methods based on the Bayesian Information Criterion (BIC) for the SPKR task.

The AMI consortium also contributed several of the newly-collected scenario meeting series to RT-05S for use as development and evaluation data, and a number of additional meetings for system

training. Special thanks are due to Jean Carletta and the annotation staff at Edinburgh for accelerating the transcription process to make this data available to the evaluation players. Providing evaluation data not only allows test sites to calibrate system performance and test robustness across a wide variety of meeting room configurations, but also provides AMI with the opportunity to help define the benchmarks for the emerging Meetings domain. Other «conference room» data providers in this year's evaluation were CMU, ICSI, and NIST, all of whom had contributed data to past evaluations, and another new contributor, Virginia Tech.

### A First Look at CHIL's Lecture Room Data

One of the novelties of this year's evaluation was the introduction of a separate «lecture room» track, targeting seminar-style meeting data collected as part of the EU FP6 Integrated Project, CHIL: Computers in the Human Interaction Loop (<http://chil.server.de>). The CHIL consortium contributed several seminar excerpts for use as development and evaluation data for the evaluation. These recordings are primarily single-speaker lectures followed by question-and-answer exchanges. Audio data was captured via elaborate multi-element microphone arrays and «inverted-T» 4-element localization arrays, in addition to the usual close-talking headset and tabletop mics.



While no additional training data was available for this new data collection, the evaluation did provide the opportunity to measure out-of-the-box system performance on an interesting new data source and did facilitate the piloting of the new SLOC task, this year evaluated only on the CHIL data.

### The Evaluation Workshop

NIST's RT-05S evaluation workshop is being held in conjunction with MLMI'05 in Edinburgh, on July 13. Evaluation players (and interested others) will come together to report on evaluation systems, compare performance, and discuss challenges in processing data from Meetings. Further details of the evaluation systems and their performance should be available on NIST's website:

<http://www.nist.gov/speech/tests/rt/rt2005/spring/> following the evaluation workshop.

## Upcoming Events

### ICMI 2005 *October 3-7, 2005*

International Conference on  
Multimodal Interfaces 2005  
<http://icmi05.itc.it>

#### Call for papers for associated workshops

- Linguistic Engineering Meets Cognitive Engineering in the Interface Design of Multimodal Systems (LECEMS)  
<http://hmi.ewi.utwente.nl/conference/LECEMS>
- Multimodal Interaction for the Visualization and Exploration of Scientific Data  
<http://www.science.uva.nl/~elenaz/ICMI/>
- International Workshop on Multimodal Multiparty Meeting Processing  
<http://www.idiap.ch/ICMI05/>
- User-centred design and evaluation of services for human-human communication and collaboration  
<http://www.Industrialdesign.tue.nl/ICMI/eval-workshop>
- Workshop on Multimodal Interaction for the Visualization and Exploration of Scientific  
<http://www.science.uva.nl/~elenaz/ICMI/>

### Journal of Machine Learning Research Special Topic on Machine Learning and Large Scale Optimizazion

#### Call for papers

We invite papers on the combination of machine learning and large scale optimization for a special topic of the Journal of Machine Learning Research (JMLR). This topic motivated the PASCAL (Pattern Analysis, Statistical Modelling and Computational Learning Network of Excellence) Workshop on «Machine Learning, SVM and Large Scale Optimization», celebrated in Thurnau, Germany from March 16 to 18, 2005.

All participants in the workshop are invited to submit a full paper, but submission is open to everyone.

Many modern machine learning algorithms reduce to solving large-scale linear, quadratic or semi-definite mathematical programming problems. Optimization has thus become a crucial tool for learning, and learning a major application of optimization. Furthermore, a systematic recasting of learning and estimation problems in the framework of mathematical programming has encouraged the use of advanced techniques from optimization such as convex analysis, Lagrangian duality and large scale linear algebra. This has allowed much sharper theoretical analyses, and greatly increased the size and range of problems that can be handled. Several key application domains have developed explosively, notably text and web analysis, machine vision, and speech all fuelled by ever expanding data resources easily

accessible via the web.

This special topic is intended to bring closer optimization and machine learning communities for further algorithmic progress, particularly for developing large-scale learning methods capable of handling massive document and image datasets.

#### Topics of interest

- Mathematical programming approaches to machine learning problems, like semi-definite programming, interior point methods, sequential convex programming, gradient-based methods, etc.
- Optimisation on graphical models for machine learning, belief propagation.
- Machines, incremental SVMs, optimization over kernels.
- Convex relaxations of machine learning problems.
- Applications involving large scale databases, such as data mining, bioinformatics, multimedia.

#### Submission Procedure

Submit papers to standard JMLR submission system, <http://jmlr.csail.mit.edu/manu/db>  
Please include a note stating that your submission is for the special topic on Machine Learning and Large Scale Optimization.

#### Important Dates

- First version due: September 15, 2005.
- First notification of acceptance or rejection: November 23, 2005
- Second version due: January 6, 2006
- Final notification of acceptance: February 15, 2006
- Final version due: March 17, 2006.

#### Guest Editors

For further details or enquires, please contact the guest editors: Kristin P. Bennett ([bennek@rpi.edu](mailto:bennek@rpi.edu)), Emilio Parrado-Hernandez ([emipar@tsc.uc3m.es](mailto:emipar@tsc.uc3m.es))

### Workshop HAREM 2005 Human Activity Recognition and Modelling September 9th, 2005, Oxford, UK

Recent advances in computer vision and learning methodologies, together with the massive increase of computational power of standard computers enabled the deployment of a new generation of computer vision systems that go beyond more traditional approaches designed mainly for modelling the scene geometry from video data. Instead, the challenge is to develop systems able to detect and signal interesting events and understand, interpret and describe the observed video sequence.

One potential application of such methodologies lies in the context of video surveillance. Due to the

massive number of cameras deployed in public spaces, it is no longer possible (or efficient) to have human operators continuously monitoring a multitude of video channels. Instead we need to provide computer vision systems able to process these video streams, use and learn contextual information, characterize the behaviour in a given situation and trigger an alert, only when an interesting «event» is detected. Needless to say, these systems must be designed to function and learn in a open-ended way and must have built-in self-regulatory, and reconfigurable capabilities.

The goal of the workshop is to bring together researchers in cognitive computer vision and particular in the domains of surveillance and human activity recognition, with an emphasis on the following topics:

- Low-level feature extraction and selection
- Modelling of attention and control
- Human activity recognition
- Cognitive and self-adaptive architectures
- Video interpretation

#### Organising Committee

Jose Santos-Victor, Instituto Superior Técnico, Portugal (Chair)  
Bob Fisher, Univ. of Edinburgh, UK  
James Crowley INRIA Rhone Alpes, France

- In conjunction with BMVC 2005 (<http://cms.brookes.ac.uk/computing/bmvc2005/>)
- Sponsored by EU Project IST-2001-37540 CAVIAR:
- Local organisation by: Instituto Superior Técnico/ Institute for Systems and Robotics <http://www.isr.ist.utl.pt/~jasv>

#### Submission Dates

Deadline Submission: June 1st, 2005  
Notification of acceptance: July 1st, 2005  
Camera-ready copy due: July 20th 2005  
Workshop September 9, 2005  
Please submit your paper as a PDF email attachment to: [jasv-at-isr.ist.utl.pt](mailto:jasv-at-isr.ist.utl.pt)

Papers should not be longer than 10 pages (use BMVC format). The paper should be kept anonymous during the review process. Any indications of name and affiliation should be removed from the paper.

In the email body, please include the following information: Title, Authors (first and last names), Affiliation of authors (institute, address), corresponding author (including his/her address, email, phone, fax)

Papers that have also been submitted to the main BMVC conference will be considered for review. Double submission must be indicated by authors and papers withdrawn if accepted for BMVC. <http://www.isr.ist.utl.pt/~jasv/harem2005>