Issue N° 7 / October 2005

Editor : Céline Aymon, celine.aymon@idiap.ch

Augmented Multi-party Interaction

AMI Newsletter

Contents

COVER STORY
The AMI Training Program 1

FOCUS
• Meet the AMI Trainees 2
• Meet the AMI Trainees 3

INSIDE AMI
• Meet the AMI Trainees 4

News

2005-09-30
The 7th issue of the AMI Newsletter is online.

www.amiproject.org

AMI c/o IDIAP Research Institute, Simplon 4, P.O. Box 592, CH-1920 Martigny
info@amiproject.org - www.amiproject.org

Cover Story

The AMI Training Program

AMI is funded by the EC an 'Integrated Project'. IPs have wider responsibilities in addition to discharging their scientific programme. One of these responsibilities is to Technology Transfer. Another is to 'Spread Excellence'. The AMI training programme *(http://www.amiproject.org/edu.php)* addresses this second goal.

AMI is an inherently multidisciplinary project. Its researchers have a variety of backgrounds, in (at least), vision, audition, language, cognition, electrical engineering, mathematics and computer science. Few, if any, educational curricula address this range of disciplines and, as a consequence, there is a shortage of researchers who can work effectively within Multimodal Interfaces. The AMI training programme is making a significant contribution to relieving this shortfall, by providing opportunities for young scientists to work within the framework of an IP.

The main activity within the training programme is the placement scheme. Students and researchers are funded to spend periods of between 3 and 18 months working in AMI labs on AMI-related projects. Each trainee has a designated supervisor in her/his host lab and is required to produce a project description at the start of the placement and a report at the end.

The placement programme is not allowed to pay salaries: what we provide is travel expenses and a living allowance of €1,250 per month. Applications are dealt with by the AMI Training Panel, which is chaired by Professor Phil Green of USFD.

AMI's US partner, the International Computer Science Institute at Berkeley, CA, has a special place in the training programme. ICSI has a long history of providing European researchers with scientific training at the leading edge of Computer Science and the AMI training programme is to some extent modelled on ICSI's experience. A separate budget funds placements at ICSI, where the living allowance is €2,000 per month, in line with the high cost of living in the Bay Area.

The placement scheme has proved popular, with 31 trainees accepted in the first two years of AMI. You can read about some of them and their projects overleaf. The scheme operates at academic levels ranging from undergraduate internships through masters and doctoral level students to post-docs. There is a roughly equal split over these levels: 7 undergradues, 8 masters, 10 doctoral and 6 post-doctoral. Eight AMI labs have acted as hosts: 11 trainees have gone to ICSI, 4 each to IDIAP, Edinburgh, Sheffield and Twente, 2 to TNO and one each to Munich and Brno. There is no nationality restriction on trainees: 8 of the 31 trainees are non- European and we have taken people from 14 countries. We are indeed 'spreading excellence': only 11 of the 31 trainees are from AMI labs.

In addition to gaining scientific knowledge and skills, we feel that AMI trainees benefit enormously from immersion into a state-of-the-art research culture. They get experience of the whole research process, from proposal to publication. They meet and interact with some of the best people in the field. They see how scientific curiosity leads to engineering solutions.

The training programme also supports Summer schools, proving funding for invited speakers and for trainees to attend. We have provided funding to the Summer schools associated with the existing Euromasters in Language and Speech and linked their Summer school with the MLMI workshop sponsored by AMI and related projects.

There is still time to apply to the AMI training programme. See *(http://www.amiproject.org/edu.php)*

Augmented Multiparty Interaction (AMI) is an Integrated Project funded by the EC's 6th Framework Program, jointly managed by IDIAP (CH) and the University of Edinburgh (UK).

Augmented Multi-party Interaction
AMI Newsletter
Issue N° 7 / October 2005
Editor : Céline Aymon, celine.aymon@idiap.ch

The AMI Training Program

MEET THE AMI TRAINEES

ICSI Visitor: Frantisek Grezl (Ph.D. Student)
Visiting From: Brno University of Technology, Czech Republic
Period: 1 Nov 2004 - 31 Mar 2005 (then continuing for 5 months at IDIAP)
Project Title: USING LONG-TERM (TRAPS-BASED) FEATURES FOR IMPROVED MEETINGS ASR

Abstract: State-of-the-art feature extraction is now moving beyond the standard simple cepstrum computation of single speech frames accompanied with their deltas. Emerging techniques involve nonlinear transformations (e.g. via neural nets), phone/state class posterior estimation, and feature combination. In addition, the signal duration for feature computation is expanding from 25 ms up to 500 ms. Front-end processing may incorporate a combination of standard short-term cepstral features plus deltas (timespan <100ms) together with either TANDEM features (timespan up to 200 ms) or TRAPS-based features (timespan up to 500 ms). The combination of these features can by as simple as their concatenation or involve more sophisticated combination, e.g. via HLDA transforms. Front-ends using simple concatenation of long-term features with standard cepstral features have recently been used with great success in automatic speech recognition (ASR) systems for transcribing conversational telephone speech, achieving relative reductions of up to 10% in word error rate. The aim of this project is to address long-term (TRAPS-based) features in the context of Meetings recognition, especially newly proposed techniques for deriving such features (HATS, TMLP), and to explore possibilities for combination of short-term and long-term features using more advanced techniques such as HLDA.

From left to right: Frantisek Grezl, Marc Ferras, Xavier Anguera and Michael Pucher

ICSI Visitor: Marc Ferras (Masters candidate)
Visiting From: Polytechnical University of Catalonia (UPC), Barcelona
Period: Started September 1, 2004, for 6 months (with possible extension to full year)
Project Title: MULTI-CHANNEL SPEECH DEREVERBERATION FOR ASR THROUGH EXPLICIT SPEECH MODELLING

Abstract: Word accuracy of ASR systems falls off dramatically when using far-field microphones, yet the use of tabletop microphones is both convenient and common in meeting room recordings. Thus, it is essential to have a preprocessing stage which copes with reverberation while trying to simultaneously maximize word accuracy of the ASR system. For this project, some already existing dereverberation techniques will be studied, implemented and evaluated for the available meeting corpora. These techniques are focused on both beamforming and speech modelling at the signal level (LPC, HNM). At the same time, other related and novel approaches will also be examined, aimed at joint beamforming-LPC (or PLP) modelling, and may involve pitch tracking or working on different metrics for LPC residual minimization.

ICSI Visitor: Michael Pucher (Ph.D. student)
Visiting From: Telecommunications Research Center, Vienna
Period: started February 1, 2005, for six months
Project Title: LATENT SEMANTIC ANALYSIS BASED LANGUAGE MODELS FOR MEETING RECOGNITION

Abstract: Language models that combine N-gram models with Latent Semantic Analysis (LSA) based models have been successfully applied for conversational speech recognition and for broadcast news. LSA defines a semantic similarity space using a large training corpus. This semantic similarity can be used for dealing with long distance dependencies, which are a problem for N-gram based models. Since LSA based models are sensitive to the topics of the training data and meetings mostly have a restricted topic or agenda, we think that these models can improve speech recognition accuracy on meetings. In this project the performance of LSA based language models on meeting recognition will be evaluated. For the training of the LSA model we will use topicalized meeting data together with larger training corpora. There are two crucial aspects of LSA based language models that we want to work on. The first is the conversion from the semantic similarity space to the probabilistic space of language models. The second is the integration of N-gram models and LSA based semantic models. We want to investigate different methods for dealing with these two issues in the meeting domain.

IDIAP Visitor: Guillaume Heusch
Visiting From: EPFL, CH
Period: Started September 1, 2004, for 6 months
Project Title: IMAGE PREPROCESSING FOR LIGHTING INVARIANT FACE RECOGNITION

Abstract: Face image processing is an important research area in AMI (Detection, Tracking and Recognition), especially in the context of meeting room data analysis. Lighting is a significant factor affecting the appearance of faces. The goal of this project is to study and to implement some state-of- the-art face image lighting normalization techniques [1,2]. As a first step, the student will study the effect of lighting change on the face recognition algorithms developed at IDIAP [3]. Then, he will study and implement the above image normalization techniques. Finally, an experimental comparison of selected techniques will be performed on a specific face recognition task. References: [1] Georghiades A., Kriegman D., Bielhumeur P., «From Few to Many: Generative Models for Recognition Under Variable Pose and Illumination», IEEE PAMI (2001) [2] Ralph Gross and Vladimir Brajovic, «An Image Preprocessing Algorithm for Illumination Invariant Face Recognition», 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), 2003 [3] F. Cardinaux, C. Sanderson, and S. Marcel, «Comparison of MLP and GMM Classifiers for Face Verification on XM2VTS, in 4th International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA, Guilford, UK, 2003, pp. 91

Guillaume Heusch

AMI c/o IDIAP Research Institute, Simplon 4, P.O. Box 592, CH-1920 Martigny, info@amiproject.org - www.amiproject.org

Augmented Multi-party Interaction

AMI Newsletter

Issue N° 7 / July 2005

Editor : Céline Aymon, celine.aymon@idiap.ch

The AMI Training Program

MEET THE AMI TRAINEES

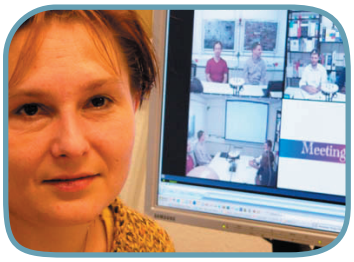
IDIAP Visitor: Hari Krishna Maganti
Visiting From: Ulm University, Germany
Period: Started October 1, 2004, for one year
Project Title: REAL-TIME UNSUPERVISED SPEAKER SEGMENTATION AND TRACKING USING SOURCE LOCALIZATION AND ACOUSTIC INFORMATION (ROBUST FEATURES)

Abstract: Speaker segmentation and tracking are crucial to the AMI project and in many applications such as in speech acquisition and recognition and meeting rooms. In the context of meeting room conversations, the speech stream is continuous and there is no information about the location of boundries between speakers- the «speaker segmentation» problem and also which portions of the speech belong to which speaker- the «speaker tracking» problem. The goal of this project is to find solutions to «who (which speaker) spoke when (at what time), where (location), and what (transcription)» using the source localization and acoustic information.

Hari Krishna Maganti

UT Visitor: Volha Petukhova (Masters candidate)
Visiting From: Tilburg University, Netherlands
Period: Started February 1, 2005, for five months
Project Title: EMPIRICALLY-BASED RESEARCH OF NON-VERBAL COMMUNICATION IN MEETINGS

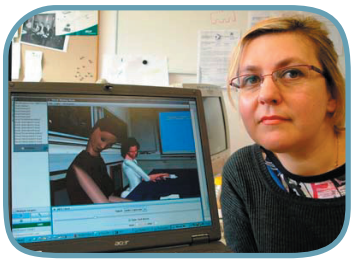
Abstract: Communication has a central place in meetings. Communication= def. transmission of content X from a sender Y to a recipient Z using an expression W and a medium Q in an environment E with a purpose/function F (Allwood, 2002). Expressions could be verbal or non-verbal in nature. The main aims of the research project are to study the interaction of verbal and non-verbal dialogue acts; to explore the semantic and pragmatic information that is available in the individual modalities; to investigate the function of non-verbal behavior, gestures in particularly; the multidimensional interaction of the verbal and non-verbal communicative acts and/or linguistic and non-linguistic components of utterances. References: Allwood, J. (2000) Bodily Communication - Dimensions of Expression and Content. Multimodality in Language and Speech Systems. Björn Granström, David House and Inger Karlsson (Eds.). Kluwer Academic Publishers. Dordrecht, The Netherlands



Volha Petukhova

Twente Visitor: Sophie-Anne Thobie
Visiting From: LIMSI-CNRS, Bordeaux
Period: 1 March - 1 September 2005
Project Title: MULTIMODAL COMMUNICATION IN COMPLEX MEETING SITUATION

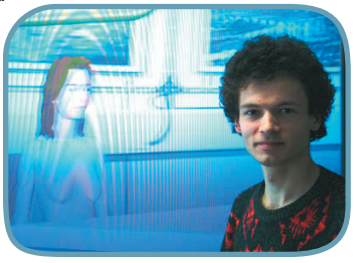
Abstract: The main feature of the work to be done in the AMI training programme is to find verbal and nonverbal characteristics of confusion during meetings. How does a meeting participant or its representation as an embodied conversational agent act when it lacks understanding of the situation? What kind of verbal and nonverbal (gaze, gestures, facial expressions, posture) show this confusion and how can we express them in embodied agents? During the traineeship an attempt will be made to model verbal and nonverbal communication issues in a situation where there is a misunderstanding among meeting participants. We will also look at the possibility to generate this type of behavior in a situation of communicating embodied conversational agents in a meeting environment. Some concrete examples of behavior, based on our model, will be generated and methods for reducing data and ameliorating the smoothness of the movements (taken from Thobie's earlier Ph.D. work) will be employed.



Sophie-Ann Thobie

Twente Visitp: Jan Peciva
Visiting From: Brno University of Technology
Period: 1 April - 23 November 2005
Project Title: MULTI-PARTY INTERACTION AND COLLABORATION BETWEEN REMOTE MEETING PARTICIPANTS

Abstract: The virtual meeting room project at the University of Twente is focused on the creation of a dynamic 3D representation of a meeting. This virtual meeting environment is meant to validate models of face-to-face verbal and nonverbal meeting behavior, but it also allows to experiment with remote participation by one or more meeting participants. In the context of the AMI design meetings it has become interesting to look at the meeting environment as an environment for collaborative work. In this traineeship the problems associated with maintaining the consistency in the environment for the various collaborating participants will be topic of concern. In the collaborative virtual meeting room that connects distributed meeting participants the meeting participants can see other participants represented by their avatars. They can see also head movements of other participants and whom they are looking at, their hand movements, and possibly all other information that will be present in a virtual meeting room. The collaborative environment should be realized through the development of methods for data sharing in time-sensitive manners, optimizing them for different network conditions, e.g. long latency or low bandwidth, and through the implementation and testing of real time interaction between users of the collaborative virtual meeting room.



Jan Peciva

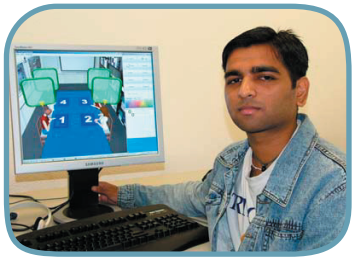
Augmented Multi-party Interaction
AMI Newsletter
Issue N° 7 / October 2005
Editor : Céline Aymon, celine.aymon@idiap.ch

The AMI Training Program

MEET THE AMI TRAINEES

Twente Visitor: Vikas Panwar
Visiting From: Indian Institute of Technology Guwahati
Period: 1 May - 31 July 2005
Project Title: DISPLAYING AND UPDATING OBJECTS DURING A DESIGN MEETING

Abstract: Interaction between participants and the objects displayed over a table is an elementry aspect of a meeting. In a virtual meeting environment, it becomes necessary that any such interaction is designed to look natural and the person who wants to discuss about the object should be able to define arguements clearly. One of the future developments that is foreseen for inclusion in this virtual meeting room is the possibility to discuss virtual objects that have been visualized in the virtual meeting environment. During design meetings these discussions include making changes or suggestions for changes for these virtual objects. Participants should be able to provide these suggestions by examining and modifying the object of attention. For such an interaction, there is a need for a suitable interface which should provide the functions or actions needed for examining or modifying the object by changing its physical properties such as size, color, position, orientation etc. Such an interface could be a combination of different modalities for example, a combination of speech and graphical user interfaces. Also,there should be an effective turn-taking process for making updates in a virtual environment. Especially to solve the problems associated with making updates and changes in the environment e.g.,in a design meeting, when several participants want to make changes to objects at the same time.



Vikas Panwar

BUT Visitor: Gaurav Pandey
Visiting from: Institute of Information Technology, Allahabad, India
Period: 26 January - 1 July 2005
Title: KEYWORD SPOTTING ON CONTINUOUS SPEECH DATA USING SEMANTIC CATEGORIES

Abstract: The work was aimed at the enhancing of acoustic (HMM-based) keyword spotting (KWS) by introducing LVCSR-loop with semantic categories to the back-ground model. First experiments aimed at learning and reproducing the results obtained with acoustic keyword spotter (the TRAP-NN-LCRC40hPostTrans system) on ICSI data. In these baseline experiments, the keyword spotting is done by traversing the data in the normal forward direction (taking left context into account). The following experiments aimed at the detection of keywords by traversing the speech data also in the reverse direction (right context). The accuracy of system in terms of figure-of-merit (FOM) was evaluated for all systems and we have tried to combine these two ways. Later on, the work was extended to include semantic categories. Firstly, by the introduction of the most common words in the free phoneme loop to add a little bit of context to the process of keyword detection. Based on satisfactory results, the work was extended to provide semantic categories. The keywords were treated according to their part of speech categories (nouns, verbs, adjectives and adverbs) and were detected using a bi-gram networks created for the different categories. In experiments, this system outperformed the baseline acoustic KWS, the improvement was especially important for the 'noun' category, where the FOM increased from 73.0% to 77.5%.

Gaurav Pandey

Robert Eklund, ICSI Visitor

Binit Mohanty, Sheffield Visitor

Robert Eklund & Matthew Aylett

Host Institute: IDIAP
- Guillaume Heusch
- Harikrishna Maganti

Host Institute: ICSI
- Frantisek Grezl
- Marc Ferràs
- Xavier Anguera
- Michael Pucher
- Mateo Agilo Bosch
- Matthew Aylett
- Rosa Martinez Torres

Host Institute: University of Twente
- Volha Petukhova
- Sophie-Anne Thobie
- Jan Peciva
- Vikas Panwar

Host Institute: TNO
- Mihaela Bobeica
- Mila Boldareva

AMI c/o IDIAP Research Institute, Simplon 4, P.O. Box 592, CH-1920 Martigny, info@amiproject.org - www.amiproject.org