FP6- 506811

AMI
AUGMENTED MULTI–PARTY INTERACTION
http://www.amiproject.org/

Integrated Project
Information Society Technologies

D.10.2 ANALYSIS OF TRAINING ACTIVITIES IN Y2 (2005)

Due date: 31/12/2005   Submission date: 4/1/2006

Project start date:1/1/2004   Duration: 36 months

Lead Contractor: USFD   Revision: 1

Table of Content

D.10.2 Analysis of training activities in Y2 (2005)

Abstract

WP10 aims to spread excellence by providing training for young researchers within the main AMI programme. The main activity is in supporting extended visits by researchers to AMI labs to work within the project's research scope. 38 such placements have so far been arranged. In addition the AMI training programme is supporting the Euromasters scheme in language and speech.

1. Introduction

The aim of the AMI training programme is to spread excellence by providing training for young researchers within the main AMI programme. The major training activity is to support placements for researchers to work in AMI labs on projects which are AMI-related. In addition, the AMI training programme provides support to the Euromasters Scheme in Language and Speech and sponsors young researchers from AMI labs to attend the Euromasters Summer School.

There is a budget of €313K for these activities over the lifetime of AMI and an additional €220K specifically to support placements at ICSI.

The AMI training programme is managed by Prof Phil Green of USFD. Prof Green chairs the AMI training panel, whose 5 members direct the programme, with the help of a part-time administrator at USFD. The training budget is handled by UEDIN.

The home page for the training programme is *http://www.amiproject.org/edu.php*.

A special AMI Newsletter (issue no. 7) featuring the AMI training programme was also released in Decmber 2005 and is available from the AMI web site at
http://www.amiproject.org/newsletter/issue07.pdf .

2. The placement programme

The AMI placement programme is available to researchers at all levels, from undergraduates to post-docs. It pays travel costs and living expenses, not salary: applicants are expected to be studying or employed at a 'home' institution. Living expenses are €1,250/month for the general scheme and €2,000/month for the ICSI scheme. Travel costs are allowed on the basis of the cheapest available flights or trains.

The placement programme is open to researchers outside the AMI labs, and to researchers outside Europe. There are no restrictions on placement length, 6 months – 1 year is typical and less than 3 months is unusual.

The application process is as follows
- Applicants apply using a form downloadable from the web site,
- They have to describe the project they want to do and why,
- They require statements of support from their home lab & their prospective host lab, together with an academic reference and CV,
- Applications are considered by the training panel on a quarterly basis,
- A budget is negotiated for each successful applicant on an individual basis,
- At this point a formal offer is made and payments are set up.

Researchers on placement are allocated supervisors at their host lab and required to produce reports at the end of the placement. Abstracts of the work in each placement are

posted on the AMI web site. At the end of a placement researchers receive an AMI Training Certificate signed by the training manager and the host institution.

The table overleaf details the placements arranged by December 31st 2005. In summary,
- 38 researchers have been offered placements, 4 have been rejected and 2 are under consideration. As might be expected, applicants 'self-select' – they are generally high-quality and well-motivated.
- The ICSI scheme has 10 takers; the other hosts are USFD(4), IDIAP(4), UEDIN(4), UT(5) , BRNO(1)& TNO(2)
- There are 7 post-docs, 12 Ph.Ds, 12 masters students and 7 undergraduates
- 11 takers are from AMI labs, 21 from outside, so the scheme really is spreading excellence
- The budget committed so far on placements is €154K for the ICSI scheme and €105K for the general scheme.

3. Other Activities

As stated in AMI's contract, we are supporting the Euromasters scheme in Language and Speech(*http://www.cstr.ed.ac.uk/emasters/*),   which   allows   students   in   participating institutes to gain an additional qualification, signed by professional bodies. To do this, they must spend at least 1 semester outside their home institution, engaged on a language/speech project. In addition, they must attend a Summer School which is organised annually. AMI gave €5K to support this Summer School in 2004 and 2005. In 2005 the Euromasters Summer School was opened to AMI researchers and tutorial material was combined.

4. Description of a few projects

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.

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

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

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 culd be verbal or non-verbal in nature. The main aims of the Volha Petukhova 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

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 fi nd 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.

Twente Visitor: 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

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 Vikas Panwar 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.

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%.

5. Status of applications (December 2005)

Applicant	Home Institute	Host Institute	Level	Visit period	Agreed cost •
Xavier Anguera Miro	Polytechnic Uni of Catalonia	ICSI	Ph.D Internship	Sep 04 - Sep 05	24,780
Marc Ferras	Polytechnic Uni of Catalonia	ICSI	Masters Internship	Start Jul-Sep End Dec-Feb	12,813
Matthew Aylett	Rhetorical Systems Ltd	ICSI	Postdoctoral Visit	April - Nov 05	16,000
Jean-Christophe Lacroix	ENST	University of Sheffield	Masters Internship	Jul-Nov 04	5,805
Guillaume Heusch	EPFL	IDIAP	U/G Internship	Sep 04 - Feb 05	7,500
Martin Karafiat	Brno	University of Sheffield	Ph.D. Internship	Jun-Sep 04	3,865
Harikrishna Maganti	University of ULM	IDIAP	Ph.D. Internship	Oct 04 - Sep 05	15,510
Michael Pucher	Telecommunications Research Center, Vienna	ICSI	Ph.D. Internship	Jan-June 05	12,594
Frantisek Grezl	Brno	ICSI/IDIAP	Ph.D. Internship	Jul 04 - Jun 05	15,461
Mihaela Bobeica	University of Nice Sophia Antipolis	TNO	Ph.D. Internship	Oct 04 - May 05	9,715
Darren Moore	IDIAP	University of Sheffield	Postdoctoral Visit	Aug 04 - Sept 04	2,830
Arlo Faria	ICSI	University of Edinburgh	U/G Internship	Feb 05 - Jul 05	8,700
Joe Frankel	CSTR Edinburgh	ICSI	Postdoctoral Visit	Sept 05 - April 06	16,000
Liudmila Boldareva	University of Twente	TNO	Ph.D Internship	Dec 04 - Dec 05	15,300
Vikas Panwar	Indian Institute of Technology Guwahati	University of Twente	U/G Internship	8 May 05 to 22 July 05	3,855
Naresh Bansal	Indian Institute of Technology Guwahati	University of Edinburgh	U/G Internship	2 May 05 to 31 July 05	4,550
Volha Petukhova	Tilburg University	University of Twente	Masters Internship	1 Feb 05 to 1 Jul 05	6,250
Robert Eklund	Linkoping University	ICSI	Postdoctoral Visit	June 05 to Feb 06	18,000

Ferran Diego	UPC Technical Uni of Catalonia	University of Edinburgh	Masters Internship	Feb 2005 to July 2005	7,900
Sophie-Anne Thobie	LIMSI/CNRS France	University of Twente	Postdoctoral Visit	1 Mar 05 to 31 Aug 05	7,860
Mateu Aguilo	University of Catalonia	ICSI	Masters Internship	Mid Feb to Mid August	12,890
Dennis Doubovski	University of Twente	Technische Universitaet Muenchen	U/G Internship	1 Feb 05 to 30 April 05	4'000
Rosa Martinez	Technical University of Catalonia	ICSI	Masters Internship	March 05 to Aug 05	12,891
Gaurav Pandey	Indian Institute of Information Technology	Brno University of Technology	U/G Internship	15 March 05 to 30 June 05	3,300
Jachym Kolar	University of West Bohemia in Pilsen	ICSI	Postdoctoral Visit	1 Oct 05 – 1 April 06	Awaiting Travel
Binit Mohanty	Indian Institute of Technology, Kanpur	University of Sheffield	U/G Internship	May 05 – July 05	4,460
Jan Peciva	Brno University of Technology	University of Twente	Ph.D Internship	1/4/05-15/5/05; 1/6/05-15/7/05; 23/8/05-23/11/05	8,280
Nynke Van der Vliet	University of Twente	University of Edinburgh	MSc Internship	1/10/05 – 30/4/06	8,825
Gerwin van Doorn	University of Twente	IDIAP	MSc Internship	1/9/05 – 31/3/06	8,875
Octavian Cheng	University of Auckland	IDIAP	PhD Internship	1/10/05 – 30/9/05 (max)	16,700
Mathias Creutz	Helsinki University of Technology	ICSI	PhD Internship	11/05 – 4/06	12,820
Cuong Huy To	IDIAP	EPFL	PhD Internship	10/05 – 8/06	PENDING
Roel Vertegaal	Queen's University	UT	Postdoctoral Visit	1/11/05 – 1/5/06	10,150
Andreu Cabrero	University Polytechnic Catalunya	UEDIN	MSc Internship	2/06 – 7/06	7,570
Anh Duc Nguyen	Vietnam National University	UT	MSc Internship	1/06 – 6/06	8,300
Anna Llagostera	Technical University of Catalonia	ICSI/UEDIN/USFD	MSc Internship	15/10/05 – 15/4/06	
Kumutha Swampillai	n/a	USFD	PhD Internship	1/1/06 – 30/6/06	
Barbara Caputo	KTH, Stockholm	IDIAP	Postdoctoral Visit	10/12/05 – 30/6/06	