Special Issue / April 2005

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

Special Issue

AMI Newsletter

Contents

Partners

Non-Profit Research Institutes
IDIAP, Research Institute (IDIAP), CH
German Research Centre for AI (DFKI), D
International Computer Science Institue, Berkley/CA (ICSI), USA
Netherlands Organisation for Applied Scientific Research (TNO), NL

Academic Partners
University of Edinburgh (UEDIN), UK
Sheffield University (USFD), UK
Brno University of Technology (BUT), CZ
Munich University of Technology (TUM), D
University of Twente (UT), NL

Industrial Partners
FastCom Technology S.A. (FC), CH
Philips Consumer Electronics BV (PHI), NL
RealVNC Ltd (VNC), UK
Spiderphone S.A (SPI), CH

Standards Representative
WWW Consortium (W3C), F

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

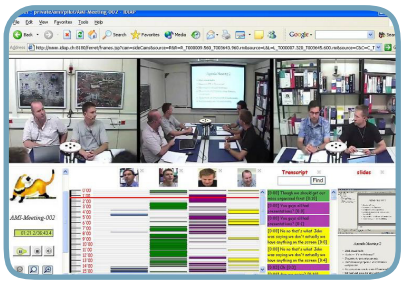
The AMI Year in Review: Establishing Foundations and Driving Technological Advances

We all have too many meetings, and often there is nothing to show except sketchy minutes or hastily written notes. The AMI (Augmented Multi-party Interaction) project aims to remedy this by providing better structure to the way meetings are run and documented. This special issue of the AMI Newsletter gives an overview of the project and describes key progress made in its first year.

AMI research revolves around instrumented meeting rooms which enable the collection, annotation, structuring, and browsing of multimodal meeting recordings. For each meeting, audio, video, slides, and textual information (notes, whiteboard text, etc) are recorded and time-synchronised. Relevant information is extracted from these raw multimodal signals using state-of-the-art processing technologies. The resulting multimedia and information streams are then available to be structured, browsed and queried within an easily accessible archive.

AMI is particularly concerned with the application of multimodal processing technologies to develop meeting browsers and remote meeting assistants. A meeting browser is a system that enables a user to navigate an archive of meetings, viewing and accessing the full multimodal content, based on automatic annotation, structuring and indexing of the information streams. For example, navigation may be enabled using automatic annotations such as speech transcription and identification of participants. A natural extension of such a meeting browser is the concept of a remote meeting assistant, which performs such operations in real time during a meeting, and enables remote participants to have a much richer interaction with the meeting.

To develop these applications, the AMI project is extending the state-of-the-art in several areas, including models of group dynamics, audio and visual processing and recognition, models to combine multiple modalities, the abstraction of content from multiparty meetings, and issues relating to human-computer interaction. These R&D themes are underpinned by the ongoing capture of user requirements, the development of a common infrastructure, and evaluations of the resultant systems. While meetings provide a rich case study for research, and a viable application market, many of the scientific advances being made within AMI supersede any single application domain. Each of the above technologies has broad application, for example in security, surveillance, home care monitoring, and in more natural human-computer interfaces.



Prototype AMI Meeting Browser

As the following pages will demonstrate, the first 12 months of the project have seen progress on several fronts: from establishing the necessary framework for successful long-term collaborative research, through to development of a prototype meeting browser. A common infrastructure has been established, and within this we are undertaking an ambitious data collection and annotation effort. We have developed state-of-the-art technology baselines in audio, visual and multimodal processing, and in content abstraction. In addition, AMI has established a series of technical workshops, an effective training programme, and has organised a successful event to foster a better understanding of the technology transfer process.

Looking to the years ahead, by the end of 2006 we expect the most significant change due to AMI to have been the furthering of multimodal processing, content abstraction and presentation technologies, enabling effective browsing of a meeting archive.

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).

Special Issue / April 2005
Editor : Céline Aymon, celine.aymon@idiap.ch
AMI Newsletter

AMI Research Themes

THIS PAGE OVERVIEWS KEY PROGRESS MADE OVER THE PAST YEAR ACROSS THE SCOPE OF THE AMI RESEARCH THEMES.

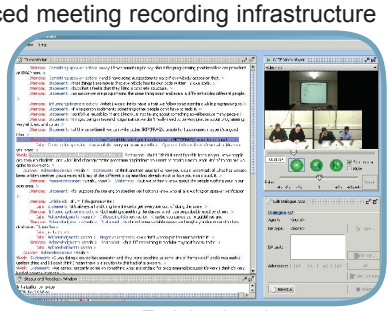
Definition and analysis of meeting scenarios

To design appropriate technologies for meetings, it is first necessary to understand the type of group, the nature of their interactions, and the means by which their members communicate. To provide a structured research framework, rather than studying a variety of unconstrained meetings, initial work in AMI is investigating scenario-based meetings. These are motivated by a scenario, or situation, which is given to participants beforehand to describe aspects relating to the meeting as a whole, such as the purpose, topic, contextual information, and expected duration. Participants in these meetings act naturally as themselves, but assume an artificial role (e.g. project manager). Use of such scenarios follows a standard methodology from social psychology, enabling the study of groups that behave as much as possible like real groups of the desired type, whilst still being able to control experimental conditions.

In AMI, a scenario has been defined based on a series of meetings in a design project. Studying a series of meetings within the context of a project allows us to not only improve understanding of processes during meetings, but also between meetings - meetings occur within a particular environment and are generally part of an ongoing work cycle. Design project meetings have the advantage of strong natural structure, e.g. progressing through phases of brainstorming, negotiation and decision-making, making them suitable for automatic techniques for structuring and summarising information. In addition, design project meetings work towards a quantifiable outcome, allowing clearer evaluation of the impact of social and organisational factors, as well as any assisting technologies being employed. To complement data collected in this scenario-driven manner, and confirm the applicability of research outcomes, other real meetings are also being collected and analysed.

Infrastructure design and data collection

Hand in hand with the scenario definition, the collection of a significant data corpus meeting the consortium's research requirements has been a major focus within the first phase of the project. Many of the technologies being developed in the project require a large amount of development data for various reasons, including rigorous testing of techniques and ensuring new techniques are robust to different operating conditions. Enhanced meeting recording infrastructure (see following page) is now operational in rooms at three AMI sites, and portable meeting recording equipment has also been developed. To date approximately 70 hours of meetings have been recorded using these facilities. This data is being annotated according to defined schemes, and distributed to partners via the AMI Media File Server, a purpose-built web server allowing researchers to view and download multimedia data relevant to their needs.



Tool developed to annotate meeting dialogue acts.

Automatic extraction of information from audio-visual data

The focus of the AMI project is to add value to meetings by automatically documenting and presenting their information content. To achieve this, the meetings are heard and observed using microphones and video cameras, and then state-of-the-art technologies are applied to extract the raw information content from these audio-visual signals. AMI technologies operating on these signals may be grouped according to six core types of information they extract: 1) recognizing what is said by participants (automatic speech recognition, see following page), 2) recognising what is done by participants (automatic action and gesture recognition), 3) recognising where each participant is at each time (localization and tracking), 4) recognising how participants act in reference to their emotional state (emotion recognition, see following page), 5) tracking what (person, object, or region) each participant is focusing on (focus-of-attention recognition), and 6) recognising the identity of each participant (person recognition).

In the first year of the project, significant progress has been made towards solving each of the above problems using state-of-the-art signal processing and machine learning technologies. Signal processing takes raw audio-visual data and reformats it in some way, e.g. extracting voice information from background noise, while machine learning techniques automatically recognize patterns present in large quantities of data, e.g. learning what a word sounds like by listening to many examples of it. Within AMI, particular focuses have been applying techniques to data recorded in real-world conditions, and using both audio and visual information whenever this is relevant and complementary (e.g. using both vocal and facial information to identify someone).

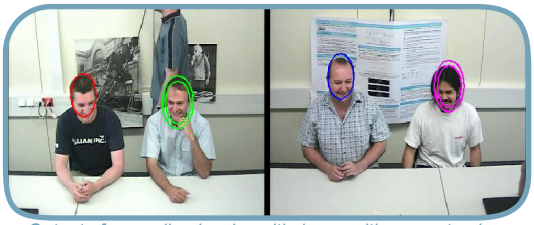
Automatic structuring and summarization of information

Meetings contain much raw information in the forms described above, such as spoken or written words, gestures, actions and emotions. To allow efficient access to relevant parts of this information, it is necessary to provide some higher-level structure or to distill the core information in the form of a summary. Within AMI, various ways of structuring meeting information are being developed, e.g. according to dialogue structure, different topic categories, and meeting phases (such as discussions or presentations). Different approaches to summarization are also being studied, including extractive summarization (in which only the key informative segments of the meeting are identified) and abstractive summarization (in which a coherent high-level text is generated to describe the most important information from the meeting).

Multimedia retrieval and presentation

The automatic information extraction technologies described above aim to supplement the raw multimedia meeting recordings with various types of meta-data – words, identities, actions, summaries, etc. A further research direction is to develop technologies that allow this multimedia data and metadata to be presented to an end-user in ways that allows them to retrieve relevant information.

Researchers in AMI are focusing on two target applications for these retrieval and presentation technologies: namely a meeting browser and a remote meeting assistant, as described on the previous page. In the first year of the project, work has focused on gathering functional user requirements for these applications, and now prototype interfaces are being developed according to these requirements. Further progress has included the implementation of a software toolkit to facilitate rapid prototyping (see following page), and development of a technique to objectively compare different user interfaces in terms of their effectiveness.



Output of an audio-visual, multi-view, multi-person tracker.

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

Special Issue / April 2005

AMI Newsletter

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

Technology Case Studies

WHILE AMI RESEARCH IS BROAD IN SCOPE, THERE IS ALSO MUCH DEPTH TO THE TECHNOLOGICAL DEVELOPMENTS IN AMI, AS TYPIFIED BY THE FOLLOWING CASE STUDIES.

The AMI Automatic Speech Recognition System

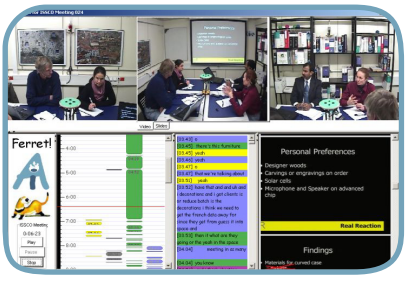
While communication in meetings is multimodal in nature, speech is the predominant mode of interaction and the richest information source. Compared to other application domains, such as dictation or broadcast news transcription, meetings pose a number of challenging problems for Automatic Speech Recognition (ASR) systems. Speech in meetings is conversational in nature and so does not follow standard grammatical constraints; AMI is addressing this by modeling the informal grammar of conversational speech, e.g. including corrections and repetitions. There are also difficulties due to multiple people talking in the same room - speaking turns can be difficult to segment and people often talk over the top of each other; by jointly processing multiple microphones AMI researchers are developing techniques to segment and separate concurrent speech. A further challenge is the occurrence of non-native speech – while English is often used in international business meetings, this is spoken with differing accents and a variable degree of fluency; AMI researchers are investigating pronunciation models that are flexible to such variations. Finally, while current ASR systems rely on headset microphones, there is a need to move towards less constraining hands-free microphones; AMI is researching the use of table-top microphone arrays which automatically track and focus sound acquisition on a particular speaker. As these new techniques emerge, they are progressively integrated into a state-of-the-art large vocabulary conversational ASR system developed by AMI researchers.

The AMI Instrumented Meeting Rooms

An important area of progress in the first year of the AMI project has been the establishment of instrumented meeting rooms to facilitate research. Facilities are operational at three sites: the University of Edinburgh, TNO Human Factors, and the IDIAP Research Institute. The rooms have been designed to allow the collection of data suitable for all the project's research needs. To ensure consistency between data across the different sites, a basic common hardware setup was adopted, including: 4 headset microphones, 2 wide-angle camera views, a tabletop unit consisting of an 8-element circular microphone array and 4 close-up cameras, a whiteboard capture device, screen capture device for a data projector, and 4 pen devices. Additional devices at different sites include lapel microphones, extra microphone arrays, other wide-angle camera views, rear-projection screens, and a binaural manikin. All data sources are acquired synchronously and at high resolutions to facilitate research of multimodal technologies. To ensure data diversity, the rooms differ in layout and wide-angle camera. In addition to this fixed infrastructure, a low-cost portable system developed at the Brno University of Technology is also available for data collection and algorithm development.

The JFerret Multimedia Browser Toolkit

JFerret is a new multimedia browser developed in the AMI project, giving researchers a common platform to develop integrated presentation demonstrators. The browser is extremely flexible, enabling almost any user interface to be composed using a combination of plug-in modules. An XML configuration file specifies which plug-in components to use, how to arrange them visually, and how they will communicate with each other. The picture here shows a sample browser implemented with JFerret – it uses 29 plug-ins, including three videos and an audio player, but is configured in less than one page of XML.



JFerret comes with a library of pre-defined plug-ins, for presentation of video, audio, slides, annotation time-lines, controls, and more. This base set of presentation components is straightforward to extend by writing new plug-ins. JFerret is written in Java, and provides a simple plug-in programming interface and an elegant communication mechanism between plug-ins. Java also allows the application to run cross-platform, either as an Applet or as a stand-alone application.

Emotion in Meetings

Complete understanding of human communication cannot be achieved without an indication of its emotional content. When a decision was taken, were people disappointed or agreeable? Were they relaxed or nervous when a certain question was asked? AMI researchers are investigating the definition of emotional content in meetings, as well as automatic techniques for its classification. To commence, a study was conducted to determine the types of emotion commonly displayed in meetings: findings included the emotional adverbs listed above, along with others such as angry or relaxed. It is however difficult to accurately categorise natural emotions, and so AMI researchers are instead adopting a dimensional approach to labeling emotion. People are continually rated along two dimensions: one axis indicates whether the emotion is negative (e.g. anger) or positive (e.g. agreeable), while the other shows if the person is exhibiting the emotion in a passive (e.g. bored) or active manner (e.g. joking). Algorithms have been developed to categorise or rate emotional content from a person's audio (e.g. speech pitch or volume) and visual cues (e.g. facial expressions), and linguistic information has also been used to improve the accuracy of automatic emotion recognition. Approaches have also been developed to recognise meeting hot-spots, which are periods when participants exhibit a high degree of involvement or interest.

Public Software Libraries

A key contribution of the AMI project to the research community is the distribution of tools and algorithms as free-source software libraries. The NITE XML Toolkit (NXT) (available at https://sourceforge.net/projects/nite/) consists of software to support human annotators and analysts working with multimodal, spoken, or text language corpora, such as the AMI Hub Corpus. The Torch machine learning library (available at http://www.torch.ch/) contains an extensive range of state-of-the-art machine learning algorithms, including neural networks, hidden Markov models, and support vector machines.

Special Issue / April 2005

AMI Newsletter

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

MLMI: Workshop on Multimodal Interaction and Related Machine Learning Algorithms

A key initiative of AMI researchers has been to establish a workshop covering many of the technical aspects of the project: specifically multimodal (human-human, or human-machine) interaction and related machine learning algorithms. The first MLMI workshop was held in Martigny, Switzerland in June 2004, and was organised in collaboration with several other related national and European projects, including M4 (http://www.m4project.org/) , IM2 (http://www.im2.ch/) and PASCAL (http://www.pascal-network.org/). The workshop attracted over 200 participants and was considered a success in terms of scientific quality, networking opportunities, and student training. Proceedings of the workshop were published as part of Springer's Lecture Notes in Computer Science and recorded oral presentations have been made available online using AMI technologies (http://mmm.idiap.ch/mlmi04/).

Building upon this success, a second MLMI workshop will be held in Edinburgh, UK, in July 2005, in conjunction with the NIST Meeting Recognition Workshop, and involving the collaboration of other related projects and networks, including CHIL (http://chil.server.de/), HUMAINE (http://emotion-research.net/) and SIMILAR (http://www.similar.cc/). Topics covered by the workshop will include:

- human-human communication modeling
- speech and visual processing
- multi-modal processing, fusion and fission
- multi-modal dialog modeling
- human-human interaction modeling
- multi-modal data structuring and presentation
- multimedia indexing and retrieval
- meeting structure analysis
- meeting summarizing
- multimodal meeting annotation
- machine learning applied to the above

MLMI 05

The AMI Training Programme

The AMI Training Programme aims to spread excellence by providing training for young researchers in domains covered by the AMI project. The major training activity is to support placements for researchers to work in AMI laboratories on AMI-related projects. The programme is available to researchers at all levels, from undergraduate to post-doctoral, and covers travel costs and living expenses. The placement programme is open to researchers outside the AMI partner laboratories, and to researchers outside Europe, and while 6-12 months is typical, there are no restrictions on placement length.

Within the first year of the programme, 16 researchers were offered placements across the AMI partners: 7 at ICSI, 3 at the University of Sheffield, 2 at the IDIAP Research Institute, 2 at TNO, 1 at the University of Edinburgh, and 1 at the University of Twente. Of these, there were 3 Post-doctoral researchers, 9 Ph.D. students, 2 Masters students and 2 undergraduates, and 8 of the researchers came from home institutions outside the AMI consortium. Full details of the placement programme are available online at http://www.amiproject.org/edu.php , along with a list of researchers sponsored by the programme and details of their projects.

In addition to the placement scheme, the AMI Training Programme provides support to the Euromasters Scheme in Language and Speech and will sponsor a Summer School for AMI researchers starting in 2005.

The AMI Technology Transfer Event

In March 2005 the AMI project organized its inaugural Technology Transfer Event, convening a group of over 50 people of various backgrounds to discuss the technology transfer process.

The workshop opened with a summary of AMI objectives and progress by Prof. Steve Renals. Over the subsequent two days, speakers included: Dr. Jun Miyazaki (Senior Principal Researcher, Fuji-Xerox FXPAL, Japan), Stan Rosenchein (President, Quindi, USA), Jim Crabtree (President, Wave3 Software, USA), Chuck House (Director, Intel Collaboratory, USA & UK), Dave Martin (Co-founder & President of SMART Technologies, Canada), Boris de Ruyter (Senior Scientist, Philips Research, NL), Christine Perey (Consultant, PEREY Consulting, CH), Bernard Gander (Vice President of Corporate Business Development, Logitech, CH), Theo Duijker (President, Arbor Audio, NL), Philipp Hoschka (Deputy Director, W3C, F), and Volker Steinbiss (CEO, Accipio Consulting, D). Speakers talked about technology transfer experiences garnered throughout their careers, covering aspects including planning, funding, productisation, and marketing. In addition to these presentations, there were active panel discussions on surmounting communications and financial barriers, a technology demonstration session, and casual networking opportunities.

The workshop was successful in establishing and renewing an active dialogue between attendees and AMI representatives, and we expect this will lead to opportunities for collaboration as the project progresses. For more information on the workshop, the list of participants and supporting materials, please visit the workshop web site at http://www.amiproject.org/ttw05/.

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