



AMI

Augmented Multiparty Interaction

Steve Renals

Centre for Speech Technology Research
University of Edinburgh

Partners



IDIAP



DFKI



TNO



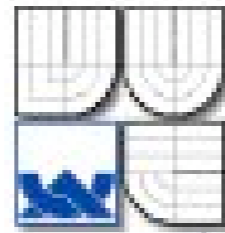
ICSI



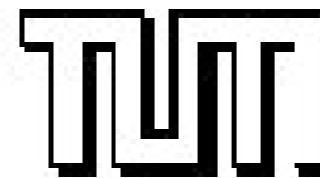
W3C



Univ of
Edinburgh



Tech Univ
of Brno



Tech Univ
of Munich



Univ of
Twente



Univ of
Sheffield



Philips



Fastcom SA



RealVnc



Spiderphone



AMI Research Vision

- Understanding human communication
 - Scene analysis
 - Unconstrained speech recognition
 - Model individuals and groups
 - Structure, index, summarize communication scenes
 - User interfaces
- Meetings provide a realistic, yet circumscribed arena to address these problems



Typical European project meeting

- **Personnel: 2 person-months**
 - Meetings: 15 people, for 15 hours.
 - Preparation: 15 people, for 5 hours.
- **Travel budget: 6000 Euros**
 - 15 airfares, at 250 E = 3750 Euro.
 - 15 nights in a hotel, at 100 E = 1500 Euro.
 - 15 dinners, 30 lunches, at 20E = 900 Euro.
 - Total = 6150 Euro.
- **Miscellaneous:**
 - 2 days back-log of other work.
 - 30 nights away from family and friends.
 - 15 Friday evenings or Saturday mornings spent in transit in London airports.

Typical results of a meeting

Former FBI Special Agent Hosty's contemporaneous
handwritten notes from November 22, 1963
post-assassination interrogation of Lee Harvey Oswald.

11/22
1026 N. Bixby room
2515 W. 5th Irving
day before yesterday
Mr. Trosky had
rifle & 2 other
1st floor outside
office
3 years in Soviet Union
relative
F.P.O.
said contacted Soviet
Embassy re wife
Hosty talking to
wife was the reason
denies Mexico City
Mrs. Paine

F.P.O. in New York
has badge for rifle
manufacture
from U.S. M.C.
O.H. Lee is ~~not~~
how he lives
but officer and he had
worked in factory
Radio Electronic Factory
dye to go home because
of confusion
1st 2nd floors offices
3, 4, 5 floors are storage
4:05 pm

Questions

- Next week
 - “What happened at the review meeting?”
- Next month
 - “Did we discuss the new tracking algorithm with the people from Munich?”
- Next year
 - “What was the precise criticism of the XYZ work?”
- Today
 - “How could the others from Maryland have participated in an efficient way?”



AMI Objectives

- Technology to support human interaction in meetings
 - Meeting Browser
 - Remote Meeting Assistant
- Based on multimodal recordings from instrumented meeting rooms
 - Audio (multiple microphones - headmounted, lapel, tabletop arrays)
 - Video (closeup and room-view)
 - Slides (data projector capture)
 - Text (handwritten notes, whiteboard)



Instrumented meeting rooms

- AMI meeting rooms at IDIAP, TNO, UEDIN
- Standardized collection of 4 person meetings using:
 - 4 close-, 2 wide-view cameras
 - 4 headset, 8 array microphones
 - data projector capture
 - whiteboard capture
 - digital pen capture
 - extra site-dependent devices (eg second microphone array, lapel mics)

Instrumented meeting rooms





Component technologies

- Browsing meetings (online and archived) requires:
 - Models of group dynamics
 - Audio and video processing and recognition
 - Models to combine modalities
 - Content extraction
 - (As well as meeting user requirements and various software technologies)
- And lots of data... well annotated



Prototype Meeting Browser

Ferret: AMI Meeting Corpus - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Print

Address C:\FerretCannedII\index.html Go Links

Whiteboard Vincent John Mark Iain

AMI Meeting Corpus

09:58:47.48.2

6:00 7:00 8:00

Rough Transcript Find

Slides Find

Project plan (1/5)

- Objective
 - Innovative TV remote control for under EUR25
- Project Budget
 - EUR 3500
- Deadline
 - At the end of this day!

Project plan (2/3)

- Budget Plan
 - Materials and equipment: EUR 300
 - People: EUR 100 per hour
 - TUR 3500-300 = 3200 : 8 hours = 4 persons
- Team
 - Project Manager
 - Industrial Designer
 - User Interface Designer
 - Marketing Repert

Project plan (3/3)

- Project Method
 - Individual work
 - Assessment stage: working
 - iterative work
 - Conceptual design meeting
 - iterative work
 - Detailed design meeting
 - iterative work



AMI data collection

- Use cases for archive browsing and online assistants
- AMI scenario meetings: set of meetings on a common design project
- Data collection in the IDIAP, TNO, Edinburgh meeting rooms
- Hub corpus: 60% scenario meetings
- Spoke corpora: ICSI and M4 corpora; specific data for localisation and tracking



Annotation

- Annotation phenomena defined - cater for all the key research problems on hub corpus
- NITE XML format and toolkit to standardize annotations
- Annotations include:
 - speech transcription
 - dialogue acts
 - focus of attention
 - summarization
 - meeting actions
 - individual actions

A Signal labeling

Continuous Signal Labeler

File Annotate View

NITE Video player
Overhead

Mute New

NITE Clock
Sync Text Areas
time: 0:07:55 skip: 5
Rate: -4x -3x -2x 0 +2x +3x +4x Reset

A - action-layer

```
[ standing up..... ] [ no action..... ] [ moving location..... ] [ no action..... ] [ moving location..... ] [ no action..... ] [ moving location..... ] [ sitting down..... ] [ no action UNFINISHED ]
```

Start: 472.4
Target: moving location
End: 478
Comment: --

Delete Edit Comment Finish

- typed e - entering
- typed l - leaving
- typed u - standing up
- typed d - sitting down
- typed m - moving location
- typed U - unconsidered action
- typed n - no action



Dialogue Act Labeling

The screenshot displays the AMI Dialogue act coder interface, which is used for labeling dialogue acts in a transcription. The interface is divided into several panels:

- Transcription:** Shows a list of dialogue turns with their corresponding labels. For example, "IS1009a.sync.4:A: Okay . Request Support: <Everybody ready > ? Uh I think the first thing we do is introduce ourselves". A context menu is open over the text "my name is Francina", listing options such as Acknowledgment, Informs, Requests, Suggests, Assessments, Social-Affective Acts, and Unclassifiable.
- Edit Adjacency Pairs:** This panel shows a list of pairs of adjacent utterances. The first pair is "A: Everybody ready" (Request Support) and "D: I think so" (Inform). The second pair is "D: I think so" (Inform) and "A: I think so" (Inform). Buttons for "Source...", "Type...", "Target...", "Set Comment", "New...", and "Delete" are visible.
- Adjacency Pairs:** A list of all adjacency pairs, including "IS1009a.adjacency-pairs.1:" and "IS1009a.adjacency-pairs.2:". The second pair is highlighted.
- Edit Dialogue Acts:** This panel shows the details of a selected dialogue act. The agent is "B", the DA type is "<none>", and the DA text is "my name is Francina". Other fields include "Addressee:" (set to "All"), "Reflexivity:" (unchecked), and buttons for "Type...", "Range...", "Set Comment", "New DA", and "Delete!".



Extractive Summarization

ICSI Extractive Summarization Coder

File Search Help

Dialogue Act Display

me018: and we do have [disfmarker]
me018: I mean ,
me018: yeah ,
me018: so [disfmarker] so you [disfmarker]
me018: yeah , it 's better to have things local if you 're gonna run over them lots of times so you don't have to go to the network .
me013: Right ,
me013: so es so especially if you 're [disfmarker] right , if you 're [disfmarker] if you 're taking some piece of the training corpus , which usually resides in where Chuck is putting it all on the [disfmarker] on the , uh , file server , uh , then , yeah , it 's fine if it 's not backed up
me013: because if it g g gets wiped out or something , y I mean it is backed up on the other disk .
me013: So ,
me018: Mm - hmm .
me013: yeah ,
me013: OK .
me018: Yeah , so , [vocalsound] one of the things that I need to [disfmarker] I 've started looking at [disfmarker]
me018: Uh , is this the appropriate time to talk about the disk space stuff ?
me013: Sure .
me018: I 've started looking at , um , disk space .
me018: Dan [disfmarker] David , um , put a new , um , drive onto Abbott , that 's an X disk ,
me018: which means it 's not backed up .
me018: So ,
me018: um , I 've been going through and copying data that is , you know , some kind of corpus stuff usually , that [disfmarker] that we 've got on a CD - ROM or something , onto that new disk to free up space [pause] on other disks .
me018: And , um , so far , um , I 've copied a couple of Carmen 's , um , databases over there .
me018: We haven't deleted them off of the slash - DC disk that they 're on right now in Abbott ,
me018: um ,
me018: uh , but we [disfmarker] I would like to go through [disfmarker] sit down with you about some of these other ones and see if we can move them onto , um , this new disk also .
mn007: Yeah , OK .

Abstract

- The main topics discussed were arrangements and objectives of an upcoming field trip to visit research partners OGI; a number of members reported their progress to date; if there are any tasks that one member can help others with; an overall description of the Cube project, a multi-lingual speech recognition system for use by the cellular phone industry, along with consideration of some of the issues therein, specifically disk and resource issues.
- Essentially the cube consists of three dimension: input features; training corpus; and test corpus.
- Most important concerns are which combinations of features to use, and what combinations of languages and broad/specific corpora to use for the training

Decisions

- The group will meet at the building at 6am to go to the airport for their field trip together.
- **Speaker me018 needs to discuss files that can be moved with speaker mn007.**
- For the OGI meeting they need to take a clear description of the cube project, and an estimate of how long the entire process should take.
- At the meeting they should discuss what they will ultimately put through the system.
- People are to consider what me034 could do on the project to speed things up, though creating the phoneme superset is a possibility.
- Speaker me018 is to look into the machines that mn007 has been running data on to find out what they are.
- Rather than consider level of normalization as a further dimension to the project, whatever OGI finds the best will be used systematically.
- Need to use multiple machines and SPERT boards to run processes on because they take so long.
- They will consider looking at articulatory features rather than straight phonemes, though it wouldn't be perfect.

Progress

- Speaker mn007 has been preparing the French digit database.
- Training and testing with varying noise.
- speaker me006 has installed updated software for everyone.
- Working on label files from TIMIT for training neural nets.
- Trying to figure out what the input to the cube should be.
- speaker fn002 has been testing the Italian database on a net trained on Spanish.
- She has had problems with incompatible labels though.
- Within the next year, the network is to be upgraded, and in a couple of weeks, the group should have access to 4 new 36 gigabyte file servers.
- **Me018 has been copying some corpus stuff to a non-backed up system, but not yet deleted originals.**
- Current plan is to use a superset of phones for the cube project derived from the various training languages.
- HTK training currently takes 6 hours to a day, and the neural net takes 1-2 days.

Monitor

8%

NITE Audio player

Synchronise
 Mute

Rate: -4x -3x -2x 0 +2x +3x +4x

Problems

- It is not clear what combinations of dimensions, which features, should be run in the cube project.
- It is important to know because the processes are going to be large and processor and memory hungry.
- To bear in mind is the fact that the cellular industry has an image of speech recognition in that's what they are after.
- Must be careful if using a broad training source that is carefully hand marked, because it would be unclear which is the reason for improvement.
- Memory is of concern, because final product needs to run potentially on cheaper cell phones, which have limited memory capacity.
- OGI doesn't have a phoneme superset ready prepared, for they are working with clusters, which may be good enough for digits, but not for discriminating words.



Media file server (mmm.idiap.ch)

AMI PILOT RECORDINGS, SERIES 1 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

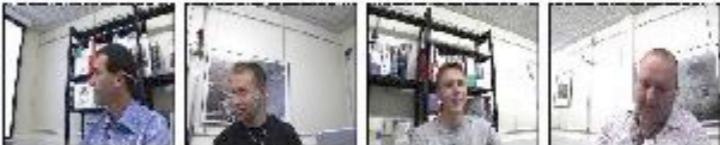

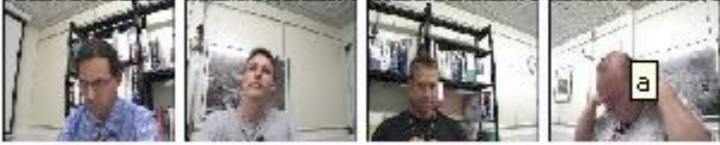
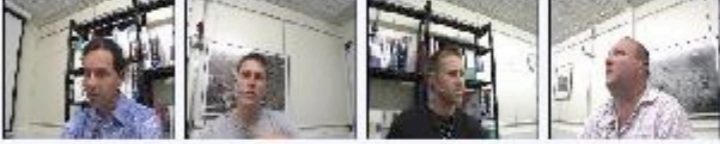
Back Forward Stop Home Search Favorites Media Print

Address <http://mmm.idiap.ch/private/AMIZone/pilot.html> Go

HELP | [MMM Home](#)

AMI PILOT RECORDINGS, SERIES 1

Please choose a meeting:

2004-08-02	AMI Meeting 001 With Ferret <i>(note: 3 seconds audio delay)</i>	 Vincent Iain John Mark
2004-08-02	AMI Meeting 002 With Ferret	 Vincent John Iain Mark
2004-08-02	AMI Meeting 003 With Ferret	 Vincent John Iain Mark
2004-08-02	AMI Meeting 004 With Ferret	 Vincent John Iain Mark

My Computer



Processing of audio-video data

- Defined according to core problems:
 - What did the participants say? (And how do they say it?)
 - What did they do? (physical actions)
 - Tracking each person's location
 - The emotional state of each participant?
 - Tracking on what each participant is focusing
 - Who are the participants?



Structuring & content extraction

- Defined according to browser requirements
 - Segmentation of multimodal streams
 - Structuring by meeting events
 - Identification of group activity
 - Indexing and retrieval
 - Summarization, and generation of textual and multimodal summaries



Multimedia presentation

- Presentation technologies for meeting data
 - JFerret browser (plugin integration framework)
 - Audio-only browser
 - New components for JFerret (eg browsing by slides)
 - Wireless presentation system
 - Virtual agents and environments for meeting playback
- Browser evaluation test

Current results

- Instrumented meeting room infrastructure
- Multimodal corpus of meeting recordings
- Meeting annotation schemes and tools
- Many component technologies: speech recognition, audio-visual tracking, summarization,
- Media file server
- JFerret meeting browser
- Open source software releases
 - NITE XML toolkit
 - TORCH machine learning toolkit