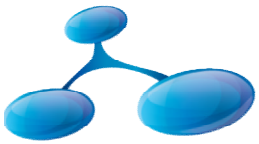
IM2 & AS Summer Institute, Riederalp, Sept. 2, 2008

Multimodal data management and annotation in the IM2 and AS NCCRs

Andrei Popescu-Belis, Martijn Goudbeek, Nele Dael, Donald Glowinski

UNIVERSITÉ DE GENÈVE



idiap RESEARCH INSTITUTE

INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

Outline of the workshop

- Schedule: 13:30 – 16:00
  - with 10' break at 14:50 (tentative)

- Objectives
  - to introduce, compare and discuss the data annotation and management techniques used in the IM2 and AS NCCRs
  - to prepare the ground for sharing these techniques in future collaborations

Schedule

13:30-13:40	*Introduction to workshop & participants*	*all*
13:40-14:10	Multimodal data management and annotation in the IM2 NCCR	A. Popescu-Belis
14:10-14:20	*Questions, discussion*	*all*
14:20-14:40	Affective sciences NCCR: database, annotation efforts and data management	M. Goudbeek
14:40-14:50	*Questions, discussion*	*all*
14:50-15:00	*Break*	
15:00-15:20	Automatic annotation of body parts and motion cues in the GEMEP database	D. Glowinski and N. Dael
15:20-15:30	*Questions, discussion*	*all*
15:30-16:00	General discussion about commonalities and differences	*all*

Multimodal data management and annotation in the IM2 NCCR

Andrei Popescu-Belis
Idiap Research Institute

IM2 & AS Summer Institute
Riederalp, September 2, 2008

Outline

1. Raw data, metadata, annotations
   - description and access

2. Explicit representation of d-m-a
   - annotations: NXT → tables
   - metadata: implicit → IMDI → tables

3. Examples of applications

Acknowledgments

- IM2 (SNSF) and AMI/AMIDA (EU) projects

- IDIAP: Mike Flynn, Bastien Crettol, Alex Nanchen
  Maël Guillemot, Pierre Wellner
- ISSCO: Paula Estrella, Philippe Baudrion
- IAM/UniBe: Marcus Liwicki, Emmanuel Indermühle
- U. of Edinburgh: Jonathan Kilgour, Jean Carletta

IM2/AMI Meeting Corpus: data sources

- Recorded meetings → 173, about 100 hours
  - 3-5 participants
  - "scenario"-based [138] = series of four meetings, the task is the design of a remote control by a group of four (with roles)
  - free-form [35]
  - recorded at three sites: IDIAP, U. of Edinburgh, TNO
- Recorded media/modalities
  - Audio (headset, lapel, array, mix)
  - Video (close up, wide angle)
  - Slides (capture)
  - Whiteboard, paper notes
  - Meeting-related documents

Metadata and annotations: definition and importance

- Metadata: static information about a meeting
- Annotations: time-dependent information from input media
  - abstraction of higher-level phenomena from low-level features
  - produced by humans ("reference") vs. extracted automatically

- Crucial role
  - reference data used for testing multimodal processing software
  - training data used for empirical research & machine learning

- Goal of the IM2 NCCR = "automation of multimodal annotation in order to enhance search and browsing"
  - IM2 = Interactive Multimodal Information Management

Main metadata items

- Participants & participant-related information
  - codename, age, gender, knowledge of English
- Date, time, location, scenario (if any)
- Relations to media-files
  - participants x channels x files
- Relations to other documents
  - produced during the meeting: slides, notes, whiteboard
  - pre- and post-meeting: agenda, supports docs, emails

Annotated phenomena: official public distribution (v1.3, 25.09.2007)

- Distribution
  - only human (= reference, ground-truth) annotations
    - except for ASR/segments output, aligned with reference
  - annotations do not always cover all meetings: too expensive
- Language and discourse
  - words, named entities, speech segments, dialogue acts
  - topics (episodes + themes), adjacency pairs
- Summaries
  - participant summaries, extractive, abstractive, links
- Argumentation: arguments & relations
- Non-linguistic modalities
  - head gesture, hand gesture, visual focus of attention, movement

Annotations in v1.3 over 173 meetings

Annotation	Meetings	%
words	171	99%
segments	171	99%
namedEntities	117	68%
dialogueActs + APs	139	80%
topics	139	80%
extractive	137	79%
abstractive	143	83%
participantSumm.	90	52%
argumentation	95	55%
headGesture	46	27%
handGesture	17	10%
focus of attention	14	8%
movement	125	72%

Other annotations

- Annotations not included in the official distribution
  - Automatic: face detection, prosody-related speech features
  - Manual: motion zones
- Annotations in progress or not yet released
  - emotion ("emotional events": positive/negative/other)
  - dominance
  - handwriting
  - BET queries: pairs of true/false statements about meeting content
- Annotations that are needed (especially "semantic" ones)
  - manual annotations → indispensable as reference
  - automatic annotations → when manual is too expensive or for comparison purposes

Example of annotation in NXT format: words and non-speech vocal sounds

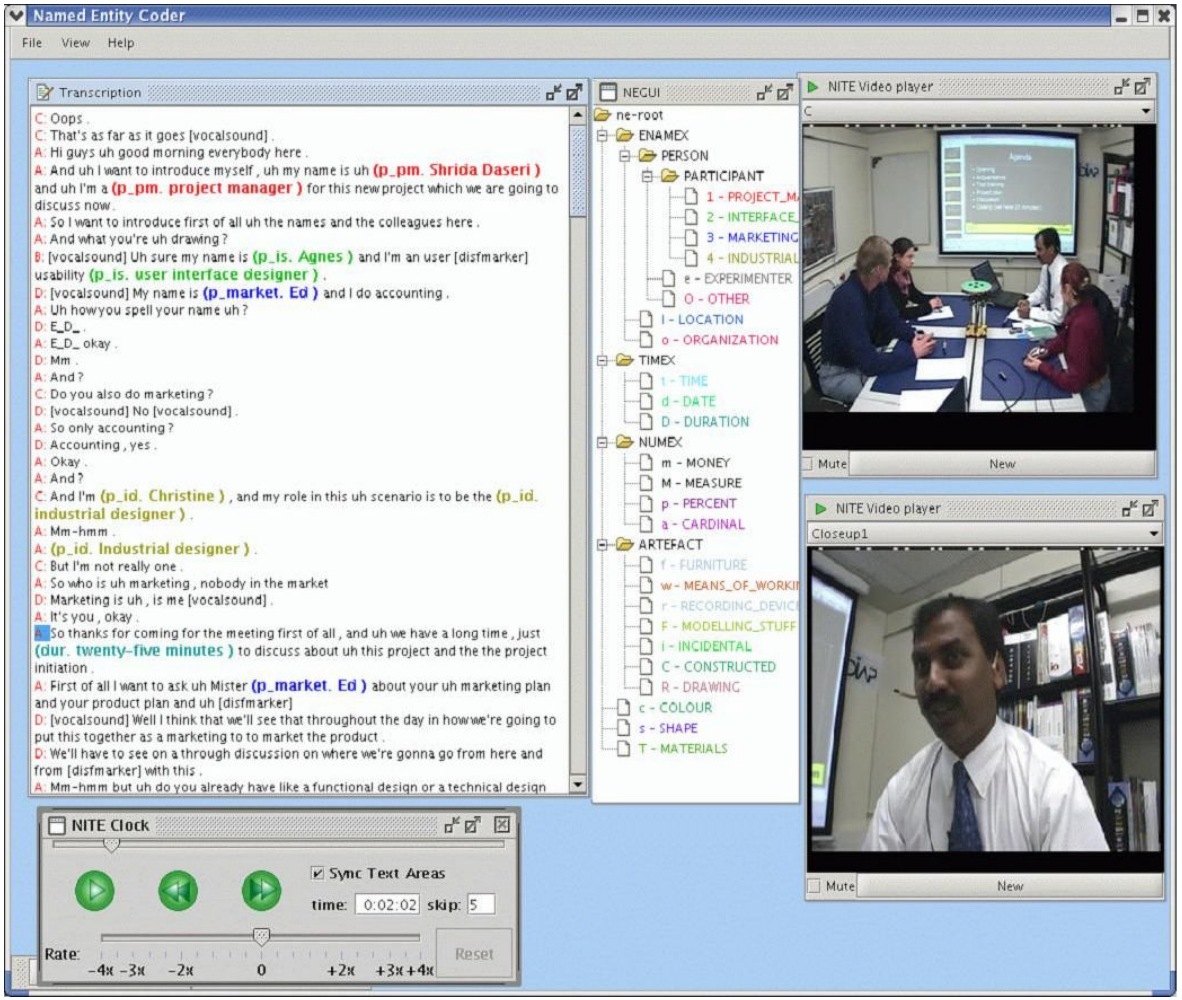
```
<nite:root nite:id="EN2001a.A.words">
    <w nite:id="EN2001a.A.words536">yet</w>
    <w nite:id="EN2001a.A.words537" punc="true">.</w>
    <vocalsound nite:id="EN2001a.A.words538" type="laugh"/>
```

- Annotation guidelines: list only 'laugh', 'other' and 'cough'
  - ~27,000 instances of vocal sounds in ~1.2 M words (2.25%)
  - the first three types: >98% of all instances
- Counts:

16508 laugh	38 sigh	20 sound indicating frustration
8919 other	29 sharp inhale	20 melodic pause filler
1114 cough	29 humming	11 sharp exhale
	20 whistling	10 pause filling noise
		9 loud exhale
		8 figurative noise
		7 singing (etc.)

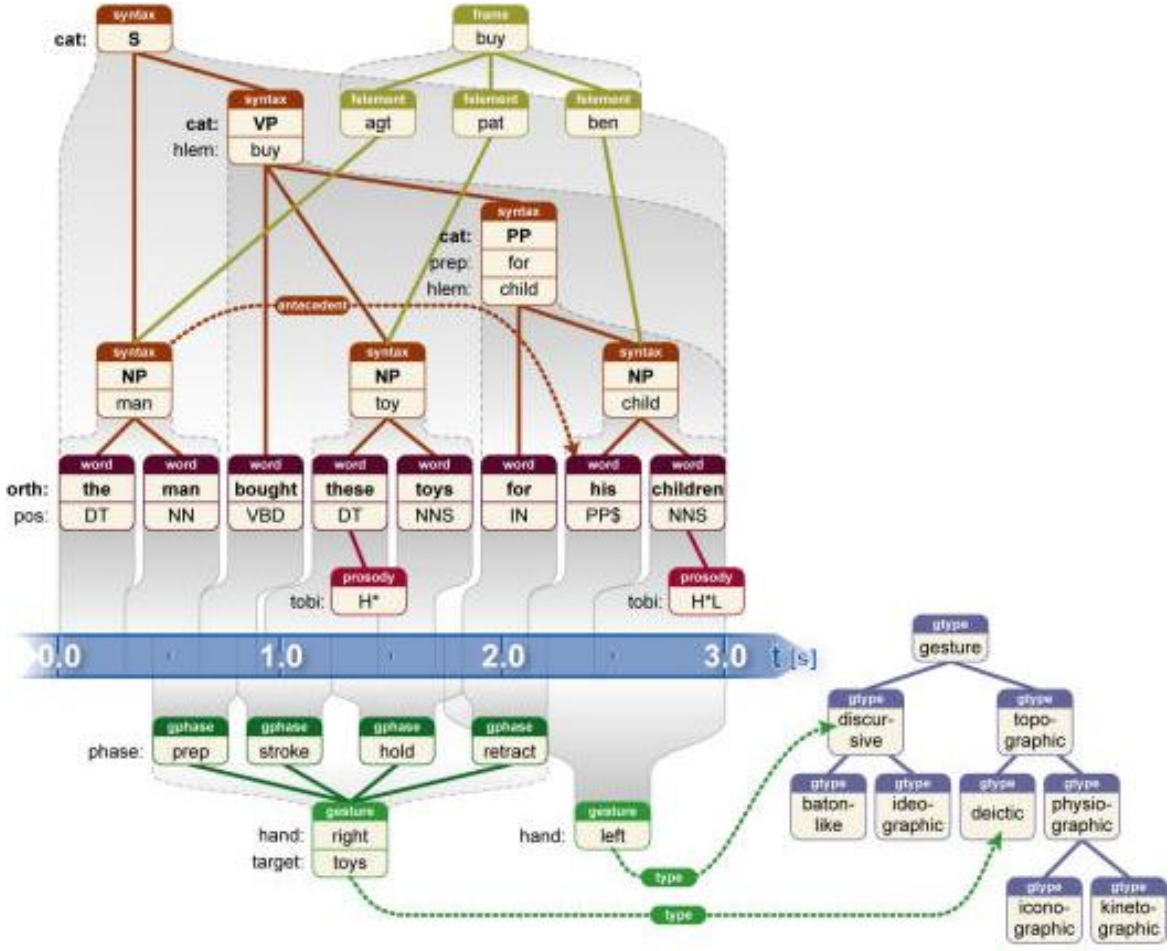
Annotation interface: NITE XML Toolkit

- Annotators do not have to manipulate the NXT format
- Toolkit can be parameterized to create interfaces for a given annotation dimension
- Text-based but with access to media files
- Most annotations of the IM2/AMI Corpus were done with NXT



NXT: a layered approach to annotation

- Time-based or text-based annotations
- Allows one to annotate upon other annotations
- Relations between annotations



Access

- http://mmm.idiap.ch
  - available corpora and annotations
    - IM2/AMI Corpus, M4, ICSI-MR, ISSCO, WSJ, Headpose, AV16 Databases, Emotional DB, IM2 Whiteboard corpus
  - AMI/IM2 Corpus
    - browse meetings | download smaller media files | download annotation archive (v. 1.3, Sep 25, 2007, 46 MB) | find documentation
    - send hard disk to IDIAP to get larger media files
- Most areas are free but require registration

navigation

- Home
- Corpora
  - AMI Corpus
  - M4 Corpus
  - ICSI Corpus
  - ISSCO Corpus
  - WSJ Corpus
  - Headpose Corpus
  - AV16 Databases
  - Emotional DB
  - IM2 Whiteboard corpus
- Tools
- Talk Webcast

AMI Corpus

— last modified 2006-08-25 16:13

The AMI Meeting Corpus is a multi-modal data set consisting of 100 hours of meeting recordings. To directly access the data, login or register here for free and go to the Access AMI corpus section. Around two-thirds of the data has been elicited using a scenario in which the participants play different roles in a design team, taking a design project from kick-off to completion over the course of a day. The rest consists of naturally occurring meetings in a range of domains. Detailed information can be found in the documentation section.

To access AMI corpus, please go to: http://corpus.amiproject.org

Synchronised recording devices:

- close-talking and far-field microphones, individual and room-view video cameras, projection, a whiteboard, individual pens.

Annotation:

- orthographic transcription, annotations for many different phenomena (dialog acts, head movement etc.).

Although the AMI Meeting Corpus was created for the uses of a consortium that is developing meeting browsing technology, it is designed to be useful for a wide range of research areas.

news

AMI annotations 1.2.2 2007-02-02

AMI annotations 1.2.1 2006-12-21

ACM UIST'06 recordings online 2006-12-20

AMI speech prosodic features 2006-11-15

IM2 Whiteboard corpus 2006-11-14

More news...

AMI CONSORTIUM

IM INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

navigation

- Home
- Corpora
  - AMI Corpus
  - M4 Corpus
  - ICSI Corpus
  - ISSCO Corpus
  - WSJ Corpus
  - Headpose Corpus
  - AV16 Databases
  - Emotional DB
  - IM2 Whiteboard corpus
- Tools
- Talk Webcast

Automatic annotations

— last modified 2006-11-15 08:34

Face Detection

The goal of face detection is to determine whether or not there are any faces in the image and, if present, their location. It is the crucial first step of any application that involves face processing systems including face recognition, face tracking, pose estimation or expression recognition.

In most applications, the face region is represented by a bounding box, such as a rectangle or a square. Here, we assume that every image is processed by a face detector and that it decides, according to specific settings, the location of faces. When a face is detected (or if there is no significant change in the image) the location and size of the bounding box (X top-left coordinate, Y top-left coordinate, Width and Height) and a confidence on the detection is returned.

PDF

Data

For more details please contact Dr. Sébastien Marcel (marcel@idiap.ch or http://www.idiap.ch/~marcel).

Speech feature outpout for prosody

Edinburgh has produced feature output for prosody for all meetings in the hub corpus. The extraction is based on individual sound files provided in the AMI corpus. To extract prosodic features from the sound files, we use the Snack Sound Toolkit to compute a list of pitch and energy values delimited by frames of 10 ms, using the normalized cross correlation function. There are 32 zip files here. The naming scheme is as ami_(energy/pitch)-ind-raw-(MEETING_COLLECTION_SITE)-(CHANNEL_ID).zip.

Download data: set of zip archives .

Main contact: Pei-yun (Sabrina) Hsueh, p.hsueh@ed.ac.uk

news

AMI annotations 1.2.2 2007-02-02

AMI annotations 1.2.1 2006-12-21

ACM UIST'06 recordings online 2006-12-20

AMI speech prosodic features 2006-11-15

IM2 Whiteboard corpus 2006-11-14

More news...

AMI CONSORTIUM

IM INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

navigation

- Home
- Access AMI corpus
- Documentation
- Sample meeting
- FAQ
- Links
- AMI corpus Contact

Welcome to the AMI Meeting Corpus

— last modified 2007-02-05 09:57

[02 Feb. 2007] AMI annotation zip archive version 1.2.2 is now available.

A DVD taster can also be ordered by completing the form (DVD Taster online order)

Register here for free

The AMI Meeting Corpus is a multi-modal data set consisting of 100 hours of meeting recordings. For a gentle introduction to the corpus, see the corpus overview. To directly access the data, login or register here for free and go to the Access AMI corpus section. Around two-thirds of the data has been elicited using a scenario in which the participants play different roles in a design team, taking a design project from kick-off to completion over the course of a day. The rest consists of naturally occurring meetings in a range of domains. Detailed information can be found in the documentation section.

Synchronised recording devices:

- close-talking and far-field microphones, individual and room-view video cameras, projection, a whiteboard, individual pens.

Annotation:

- orthographic transcription, annotations for many different phenomena (dialog acts, head movement etc.).

Although the AMI Meeting Corpus was created for the uses of a consortium that is developing meeting browsing technology, it is designed to be useful for a wide range of research areas.

SOME RIGHTS RESERVED

All of the signals and transcription, and some of the annotations, have been released publicly under the AMI Meeting Corpus license, very similar to the Creative Commons Attribution NonCommercial ShareAlike 2.5 License. The remaining annotations generated for the original project will be released in stages by January 2007, but annotations may arise out of the ShareAlike licensing for some time to come.

Copyright © 2007 by AMI project, All Rights Reserved.

Download — AMI corpus - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://corpus.amiproject.org/download/download

Google

navigation

- Home
- Access AMI corpus
  - Browse online
  - Download
  - forum
- Documentation
- Sample meeting
- FAQ
- Links
- AMI corpus Contact

Download

— last modified 2007-02-05 09:56

Use this page to download signals and annotations from the AMI corpus. The annotations, which include the orthographic transcription, come all together in one zip file. The signals are too large to package in this way, so you need to use the chooser to indicate which ones you wish to download. Full-size videos are not available from this page; for them, you need to order a fire-wire drive distribution of the data using the contact page.

Choose a mirror

Switzerland (default) | UK (to be set) | Czech Republic (to be set)

Annotations, including transcription

- AMI annotation ZIP archive (21MB, 02-02-2006), in NXT format. release version 1.2.2. To use with Signals downloaded below, unzip this file into the 'amicorpus' directory. Requires NXT version 1.3.6.

The documentation gives a release schedule for further annotations.

Signals

1) selection of one or more AMI meetings

NOTE: For scenario meetings, 1 day-recording session is divided into four [a, b, c, d] 1-hour meetings. Selecting **ES2008** meeting session together with **'a'** below allows you to get signals for **ES2008a** meeting.

Scenario Meetings

Edinburgh
ES2002
ES2003
ES2004
ES2005
ES2006
ES2007
ES2008
ES2009
ES2010
ES2011
ES2012
ES2013
ES2014
ES2015
ES2016
a b c d

IDIAP
IS1000
IS1001
IS1002
IS1003
IS1004
IS1005
IS1006
IS1007
IS1008
IS1009
a b c d

TNO
TS3003
TS3004
TS3005
TS3006
TS3007
TS3008
TS3009
TS3010
TS3011
TS3012
a b c d

Non Scenario Meetings

Edinburgh
EN2001a
EN2001b
EN2001d
EN2001e
EN2002a
EN2002b
EN2002c
EN2002d
EN2003a
EN2004a
EN2005a
EN2006a
EN2006b
EN2009b
EN2009c
EN2009d

ISSCO-IDIAP
IB4001
IB4002
IB4003
IB4004
IB4005
IB4010
IB4011

IDIAP
IN1001
IN1002
IN1005
IN1007
IN1008
IN1009
IN1012
IN1013
IN1014
IN1016

NOTE: For scenario meetings, 1 day-recording session is divided into four [a, b, c, d] 1-hour meetings. Selecting **ES2008** meeting session together with **'a'** below allows you to get signals for **ES2008a** meeting.

Terminé

Outline of the talk

1. Raw data, metadata, annotations
   - description and access

2. Explicit representation of d-m-a
   - annotations: NXT → tables
   - metadata: implicit → IMDI → tables

3. Examples of applications

Requirements for metadata and annotation database

- Informativeness
  - preserve all the relevant information from annotations
- Simplicity
  - define table structure that is easy to understand
    - one table per annotation dimension
    - other solutions: one table with triples, entity-relationship model
  - replicate some information to make tables more intelligible
  - resolve NXT pointers, especially for words
  - advantage: speed up frequent queries, especially multiple joins
- Reusability
  - easy to recreate the tables from the official annotation distribution
  - easy to modify tools if a different structure of the tables is desired

A solution for sharing annotations

1. Start with the official NXT or XML release of the annotations as a reference version
2. Apply table generation mechanism to XML annotation files, using XSL Transformations called by a script
3. Get tabular files (TSV) + table-creation script (db_loader)
4. Create and populate annotation DB
5. Customize tools (2) as needed for specific table formats

Results

- All public annotations (v1.2.2) were processed (11 MB zipped)
- Released 16 stylesheets + application script + 16 TSV files + SQL table-creation script (generated by application script)

Entries	Annotation dimension
1,207,769	words.tsv
14,230	namedEntities.tsv
69,258	segments.tsv
1,879	topics.tsv
117,043	dialogueActs.tsv
26,825	adjacencyPairs.tsv
2,578	abstractive.tsv
19,216	extractive.tsv
22,101	summlink.tsv
3,409	participantSummaries.tsv

Entries	Annotation dimension
31,271	focus.tsv
1,453	handGesture.tsv
36,257	headGesture.tsv
4,759	argumentationrels.tsv
6,920	argumentstructs.tsv
8,637	discussions.tsv

Metadata: IMDI XML-based format

- Dublin Core Metadata Initiative
  - librarians exchanging information → 15 'descriptors'
- Open Archives Initiative (OAI)
  - protocol for metadata 'harvesting' (OAI-PMH)
  - metadata providers (e.g. arXiv.org) → service providers

  → network of archives with open metadata
- For language-based resources
  - Open Language Archives Community (OLAC), extending DC
  - ISLE Metadata Intiative (IMDI): multimedia recordings

  → rich metadata file format, plus tools to edit/browse

Application to AMI/IM2 Corpus

- Static IMDI.XML metadata files
  - increase *internal* accessibility

    → no need to parse file names / analyze folders
  - increase *external* visibility

    → public catalogs with OAI-PMH

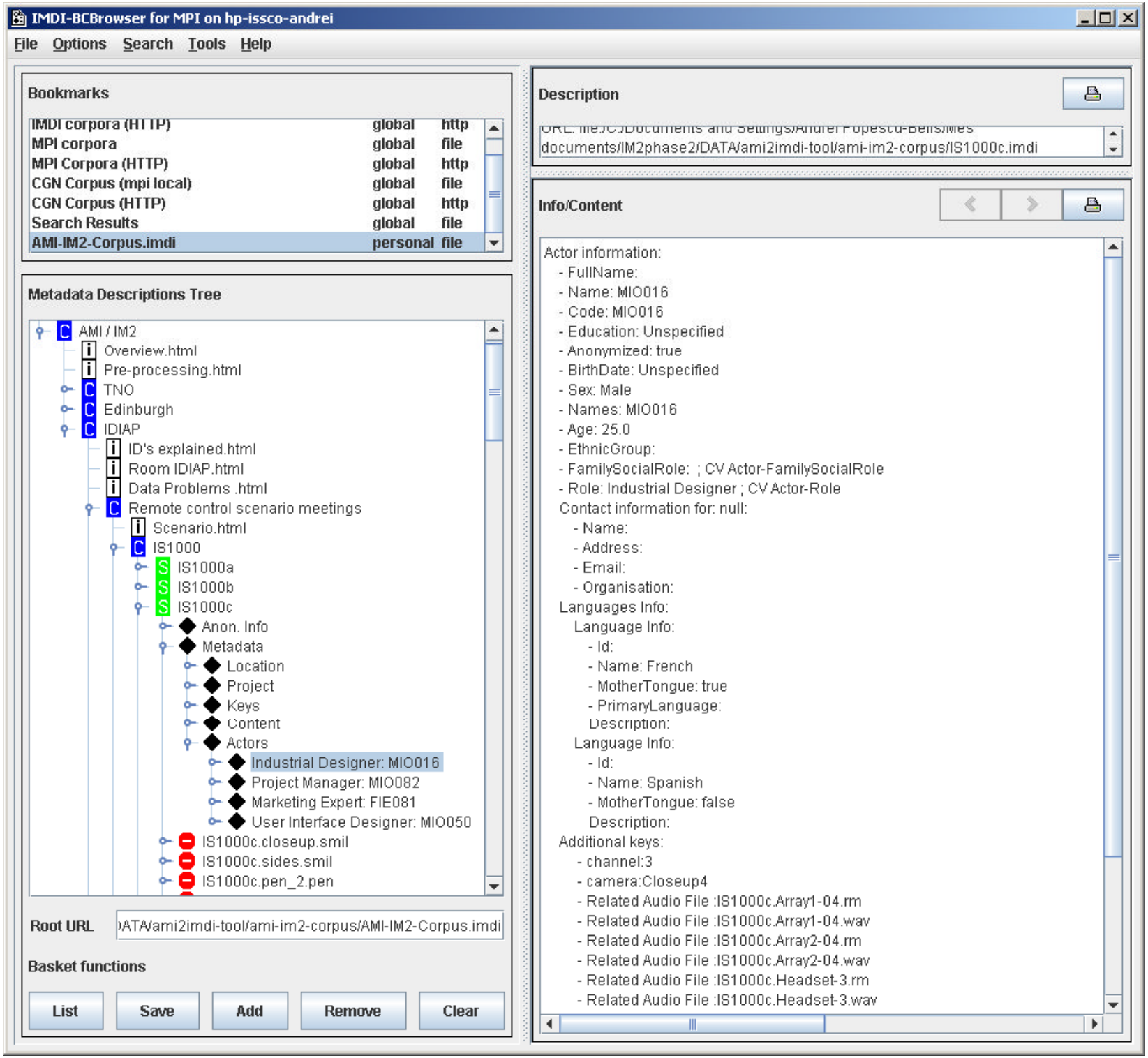
- Tools provided by IMDI to manipulate metadata
  - IMDI Editor: input manually metadata
  - IMDI Browser: visualize metadata from XML files
    - interesting feature: search metadata set

A solution for sharing metadata

1. Start with the file structure that implicitly codes metadata + two "corpus resource" files from the official release
2. Use crawling script to gather all metadata into one XML file
3. Apply an XSLT file to generate IMDI.XML declarative metadata files for each meeting
4. Apply table generation mechanism to each IMDI.XML metadata file, using second XSLT file + scripts
5. Get tabular files (TSV) and table-creation script (db_loader)
6. Create and populate annotation database
7. Customize tools (2) as needed for specific table formats

Results

- Gathered file info from MMM into a unique XML file
- Parsed XML file and resource file names
- Generated one IMDI.XML file per meeting
- Manually added files describing corpus structure
- Converted IMDI.XML files to DB format
  - Meetings (id, place, date, start, end, duration, type, description)
  - Participation (id, speaker_id, role, camera, channel)
  - Speakers (speaker_id, sex, age, native_lang, eng_region, eng_country, eng_months)
  - Influence_english (speaker_id, language)
  - Media files (meeting_id, resource_id, uri, format, media, size)


IMDI-BCBrowser for MPI on hp-issco-andrei
File Options Search Tools Help
Bookmarks
IMDI corpora (HTTP) global http
MPI corpora global file
MPI Corpora (HTTP) global http
CGN Corpus (mpi local) global file
CGN Corpus (HTTP) global http
Search Results global file
AMI-IM2-Corpus.imdi personal file
Metadata Descriptions Tree
AMI / IM2
Overview.html
Pre-processing.html
TNO
Edinburgh
IDIAP
ID's explained.html
Room IDIAP.html
Data Problems .html
Remote control scenario meetings
Scenario.html
IS1000
IS1000a
IS1000b
IS1000c
Anon. Info
Metadata
Location
Project
Keys
Content
Actors
Industrial Designer: MIO016
Project Manager: MIO082
Marketing Expert: FIE081
User Interface Designer: MIO050
IS1000c.closeup.smil
IS1000c.sides.smil
IS1000c.pen_2.pen
Root URL
)ATA/ami2imdi-tool/ami-im2-corpus/AMI-IM2-Corpus.imdi
Basket functions
List
Save
Add
Remove
Clear
Description
documents/IM2phase2/DATA/ami2imdi-tool/ami-im2-corpus/IS1000c.imdi
Info/Content
Actor information:
- FullName:
- Name: MIO016
- Code: MIO016
- Education: Unspecified
- Anonymized: true
- BirthDate: Unspecified
- Sex: Male
- Names: MIO016
- Age: 25.0
- EthnicGroup:
- FamilySocialRole: ; CV Actor-FamilySocialRole
- Role: Industrial Designer ; CV Actor-Role
Contact information for: null:
- Name:
- Address:
- Email:
- Organisation:
Languages Info:
Language Info:
- Id:
- Name: French
- MotherTongue: true
- PrimaryLanguage:
Description:
Language Info:
- Id:
- Name: Spanish
- MotherTongue: false
Description:
Additional keys:
- channel:3
- camera:Closeup4
- Related Audio File :IS1000c.Array1-04.rm
- Related Audio File :IS1000c.Array1-04.wav
- Related Audio File :IS1000c.Array2-04.rm
- Related Audio File :IS1000c.Array2-04.wav
- Related Audio File :IS1000c.Headset-3.rm
- Related Audio File :IS1000c.Headset-3.wav

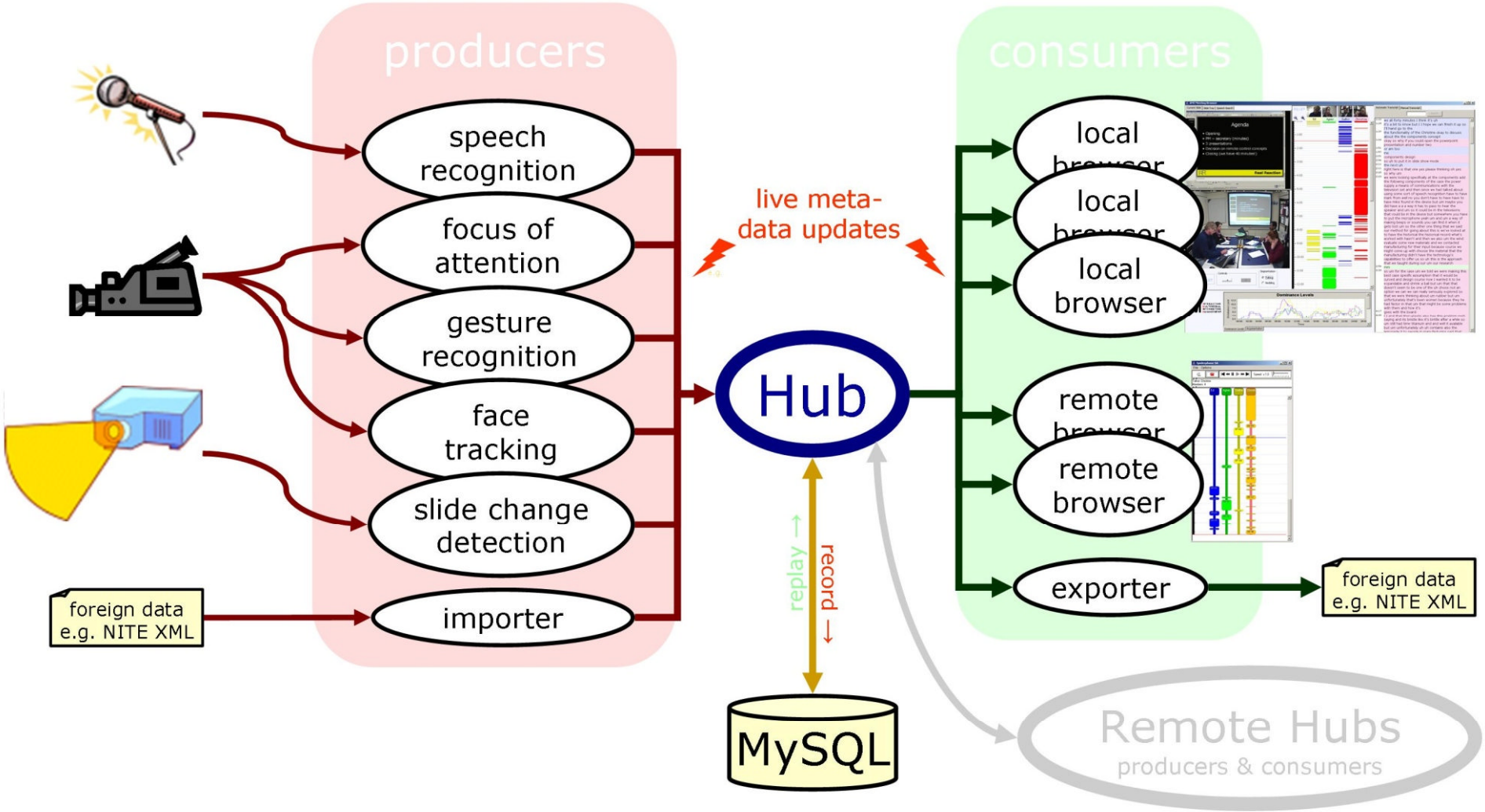
Discussion

- Advantages
  - easier to understand & use than NXT files
  - consistency checking of annotations
  - flexible, expandable
    - when needed, adjust/extend XSLT & regenerate DB
    - NXT annotations and IMDI.XML metadata serve as a reference
    - acceptable if reference does not change very often

- But: static, applicable to recorded corpora
  - dynamic solution = **The Hub**, a client-server architecture for exchanging metadata and annotations

The Hub: real-time data exchange

(borrowed from M. Flynn, B. Crettol, M. Guillemot)



➔ benefits: real-time, distributed, subscription-based, extensible

Outline of the talk

1. Raw data, metadata, annotations
   - description and access

2. Explicit representation of d-m-a
   - annotations: NXT → tables
   - metadata: implicit → IMDI → tables

3. Examples of applications

Examples of applications

- Speech segmentation, visual focus of attention, dominance
  - IM2 head pose data used in CLEAR 2007 evaluations
- Audio-visual speech recognition
- Social analysis of media
- Many others…
  - segmentation & gesture help to compute meeting actions
  - transcripts used for language processing tasks
  - dialogue acts used for summarization, decision detection
  - used in NIST Rich Transcription evaluations (2006, 2007)
- Annotations that enhance meeting browsing

Task-based evaluation [BET] of a meeting browser [TQB]

- Transcript-based Query & Browsing interface
  - intra-meeting browser
  - direct access to data-metadata-annotations
    - audio, documents, image of room, manual transcript // audio
    - annotation database (excerpts): words, utterances, dialogue acts, topic segmentation, references to documents

- Task-based evaluation: BET protocol

  → how do subjects use the browser to answer true/false questions about meetings?
    - lists of questions for 3 meetings are available on MMM

  → what are the most useful linguistic annotations?

- See my poster!

Conclusion and perspectives

- Meeting browsers
  - improve analysis of human scores
  - use annotations for automatic processing

- Data, metadata and annotations
  - improve distribution, standardization & visibility
    - include metadata in IMDI/OLAC network of providers
  - develop and/or test processing algorithms and browsers using existing annotations as a reference / training data

    → **benchmark data for multimodal processing**