

RT07: CHIL DATA

Susanne Burger
CMU

Outline

- Technical Setup
- Scenario
- Speakers
- Transcription
- DEV Set / Test Set
- Data Features

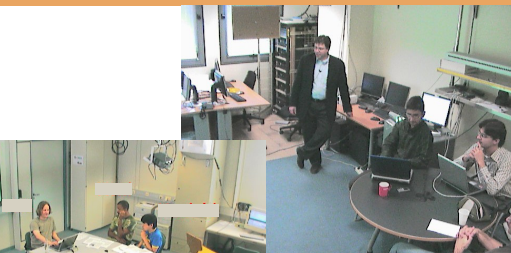
- Multi-sensory audiovisual recordings inside smart rooms (CHIL rooms)
- Five different recording sites with similar setup:
 - Variability through differences in size and layout
 - Homogenous data through common hardware and software setup

-  **AIT:** Research and Education Society in Information Technologies at Athens Information Technology, Athens, Greece
-  **IBM:** IBM T.J. Watson Research Center, Yorktown Heights, USA
-  **ITC-irst:** Centro per la ricerca scientifica e tecnologica at the Istituto Trentino di Cultura, Trento, Italy
-  **UKA:** Interactive Systems Labs of the Universität Karlsruhe, Germany
-  **UPC:** Universitat Politècnica de Catalunya, Barcelona, Spain



UPC - Spain

IBM - USA



AIT - Greece

ITC - Italy



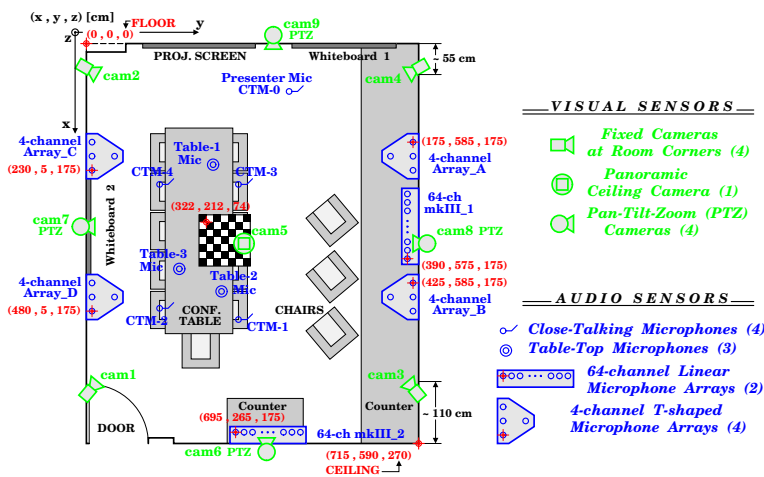
UKA - Germany



- Common sensors video:
 - Four fixed cameras located at the room corners
 - One fixed, wide-angle panoramic camera
 - One active pan-tilt-zoom camera

Mics	Mic Type	Location	Amplification	Recording	Format
64	NIST Mark III array	Wall opposite to speakers		Ethernet IP packages	SPHERE, 44.1 kHz, 28 bits
12	4-channel T-shape array	Walls			
3	table-top mics	Table	RME Octamic 8 channel,	Hammerfall HDSP9652 I/O	WAV, 44.1 kHz, 28 bits
1	CTM wireless	Presenter			
2 or more	CTM	Other participants			

Example: IBM CHIL room setup



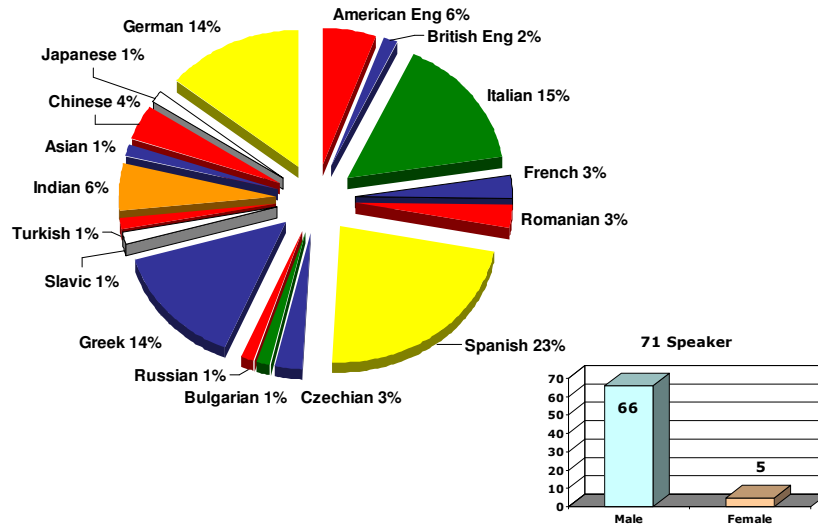
Interactive Seminars:

- 3-5 people sit around a table
- one participant presents
- other participants interrupt with questions and comments
- sometimes this results in discussions

What is "scripted" in the CHIL seminars?

- Coffee breaks
- Set of artificial background noises (cell phone, keys, phone rings, typing ..)
- Somebody enters or leaves during the meeting

But actual conversations are NOT scripted.



Transcription

- CTM channels: orthographic word level incl. vocal noise (laughter, filled pauses etc)
- Manual segmentation of talk spurts: each speaker contribution at least 300 ms away from following contribution.

- Transcription of far-field (FF) condition based on one of the table microphones
- Bootstrapped from CT transcription:
 - Non-audible elements removed, non-identifiable elements tagged, details recorded by only the FF mics added

The screenshot displays the TransEdit application window. The title bar reads "ait_20060728_ctm.trf - TransEdit 1.3 beta 4 - (C)1999-2006 by U. Mezer & S. Burger". The interface includes a menu bar (File, Edit, Labels, options, Advanced, View, Help), a toolbar, and a format dropdown set to "CHU".

Below the toolbar is a table of labels for transcription:

AAIT	BAIT	CAIT	DAIT
<Cough>	<Laugh>	<Throat>	<Smack>
<Swallow>	<Noise>	<uh>	<uhm>
<hep>	<hmo>		<P>
<T>	<FDR>	word	<T>

The main text area contains a transcript with the following entries:


```

a001_1_0080_AAIT_00: the first was information , now you have to have knowledge . I see Aristodemos has a question .
a001_3_0081_CAIT_00: */agriculture/* */al/* agriculture is <uh> out of the question , <Laugh> now , completely .
a001_1_0082_AAIT_00: completely . yes .
a001_1_0083_AAIT_00: end% <T>
a001_3_0084_CAIT_00: so I should sell */all/* <uh> all my property in Crete ?
a001_1_0085_AAIT_00: no . <P> build it so that it becomes the basis for knowledge-based applications .
a001_3_0086_CAIT_00: %*/will/* I will put some *smartness in my olive trees .
a001_4_0087_DAIT_00: <Laugh>
a001_1_0088_AAIT_00: no .
a001_1_0089_AAIT_00: <uh> make it into buildings .
a001_1_0090_AAIT_00: <Laugh> so that knowledge-based workers can come and work for you .
a001_3_0091_CAIT_00: <Noise>
    
```

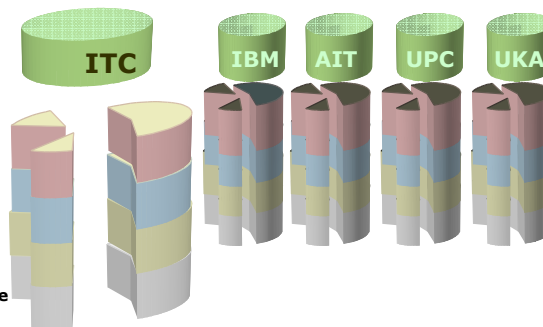
At the bottom of the window is a waveform visualization showing audio signals for several segments, with labels like "CAIT_0086_3", "DAIT_0087_4", "AAIT_0088_1", "CAIT_0089_3", and "AAIT_0090_1" overlaid on the waveforms.

...
ait_20060728_ctm_ctm_1_ait_20060728_ctm-ait_005 260.396 264.8 <o,male> the first was information , now you have to have knowledge . I see Aristodemos has a question .
ait_20060728_ctm_ctm_3_ait_20060728_ctm-ait_004 263.914 268.275 <o,male> +/agriculture/+ +/al/=+ agriculture is <uh> out of the question , <Laugh> now , completely .
ait_20060728_ctm_ctm_1_inter_segment_gap 264.8 268.196 <o,>
ait_20060728_ctm_ctm_1_ait_20060728_ctm-ait_005 268.196 269.199 <o,male> completely , yes .
ait_20060728_ctm_ctm_3_inter_segment_gap 268.275 269.69 <o,>
ait_20060728_ctm_ctm_1_inter_segment_gap 269.199 269.549 <o,>
ait_20060728_ctm_ctm_1_ait_20060728_ctm-ait_005 269.549 269.947 <o,male> and% <*T>t
ait_20060728_ctm_ctm_3_ait_20060728_ctm-ait_004 269.69 273.335 <o,male> so I should sell +/all/+ <uh> all my property in Crete ?
ait_20060728_ctm_ctm_1_inter_segment_gap 269.947 273.022 <o,>
ait_20060728_ctm_ctm_1_ait_20060728_ctm-ait_005 273.022 277.697 <o,male> no . <P> build it so that it becomes the basis for knowledge-based applications .
ait_20060728_ctm_ctm_3_inter_segment_gap 273.335 276.206 <o,>
ait_20060728_ctm_ctm_3_ait_20060728_ctm-ait_004 276.206 279.724 <o,male> so% +/I will/+ I will put some *smartness in my olive trees .
ait_20060728_ctm_ctm_1_inter_segment_gap 277.697 279.825 <o,>
ait_20060728_ctm_ctm_4_ait_20060728_ctm-ait_003 279.623 280.631 <o,male> <Laugh>
ait_20060728_ctm_ctm_3_inter_segment_gap 279.724 282.991 <o,>
ait_20060728_ctm_ctm_1_ait_20060728_ctm-ait_005 279.825 280.281 <o,male> no .
ait_20060728_ctm_ctm_1_inter_segment_gap 280.281 280.594 <o,>
ait_20060728_ctm_ctm_1_ait_20060728_ctm-ait_005 280.594 282.127 <o,male> <uh> make it into buildings .
ait_20060728_ctm_ctm_4_inter_segment_gap 280.631 333.809 <o,>
ait_20060728_ctm_ctm_1_inter_segment_gap 282.127 282.578 <o,>
ait_20060728_ctm_ctm_1_ait_20060728_ctm-ait_005 282.578 285.905 <o,male> <Laugh> so that knowledge-based workers can come and work for you .
...

5 seminars per CHIL room site, 2006 data collection:

DEV SET
1 complete seminar
per site
Total of 2 h 45 min

TEST SET
two 5 min cut-outs
of 4 seminars per site
Total of 3 h 25 min



Training data:
CHIL 2005 DEV and TEST
CHIL 2006 DEV and TEST

Selection of Test_Set segments:

1. Beginning: entering and set up
2. Meeting: presentations
3. Coffee break: real coffee breaks or people moving around.
4. Question answers: mainly at the end of meetings
5. End: last 5 min of the meeting

Each category is represented at least one time per site in similar distribution for all 40 segments

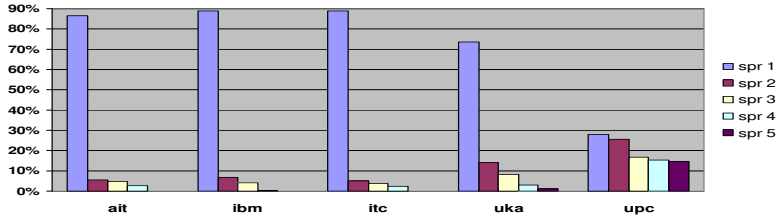
DEV_Set:

- 6656 turns
- 44300 tokens
- 2729 types
- 10% of types are proper names

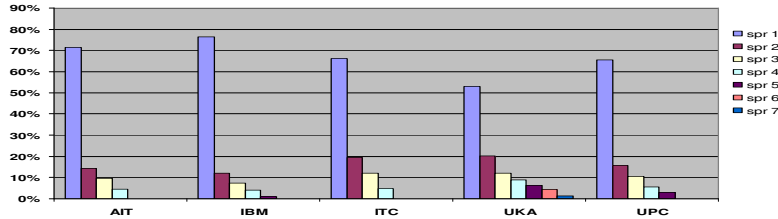
Test_Set:

- 11794 turns
- 56196 tokens
- 2870 types
- 9.5% of types are proper names

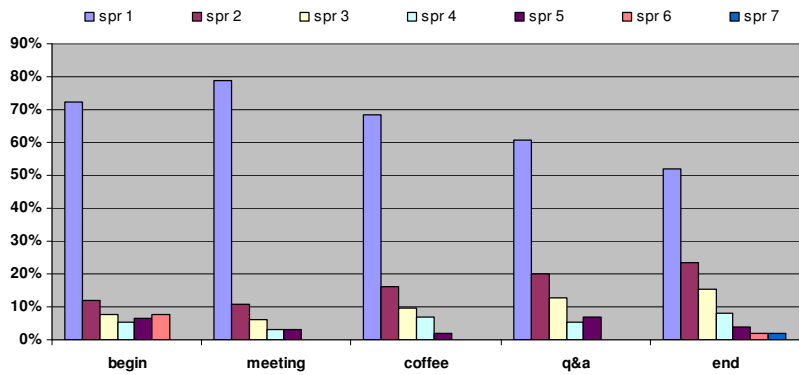
DEV Set: Average Talk-time distribution per speaker and location

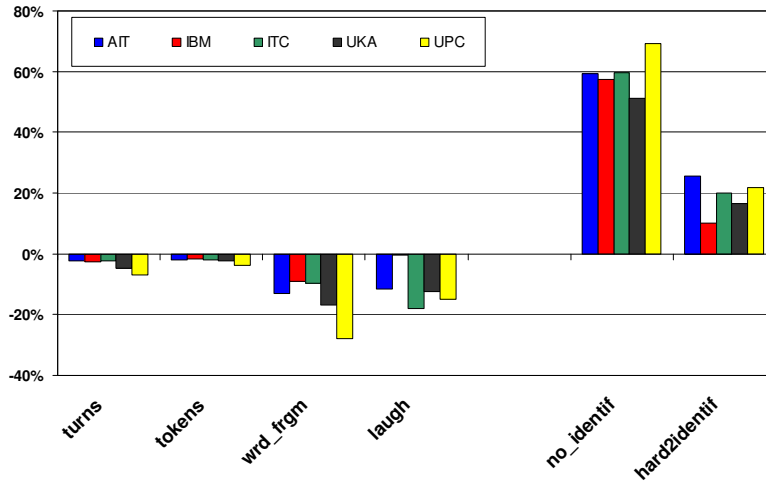


TEST Set: Average Talk-time distribution per speaker and location



TEST Set: Average Talk-time distribution per speaker and 5 min segment position





Percentage of lost / gain from CTM to FF transcription