

COCOSDA Meeting 2010, Valetta, Malta
Language Technology issues for International Cooperation

MULTIMODAL CORPORA

nick campbell
trinity college dublin

A GROWING TREND

- many corpora are now combining audio & video sources
- spoken dialogue systems will soon be incorporating cameras
- face-detection is now a standard reliable plugin for video
- motion capture devices are increasingly being used
- not just for sign-language, but for everyday conversational interaction in both social and business environments

MM-INTERACTION

- human interaction is usually multimodal -
- audio-only telephones suffice, but compare TV & radio
- facial expression and gesture signal important information
- movements and bodily alignment/posture
- making full use of redundancy in spoken interactions
- generation of AV content as well as dialogue systems

LREC MULTIMODAL

- there was a one-day workshop (tuesday)
- and 33 papers presented in 3 sessions (5%)
- in the world:
 - AMI/AMIDA might be the best known large project

ONE ORAL, TWO POSTER & ONE WORKSHOP

- Tuesday 18th - W6 -> Multimodal Corpora - Advances in Capturing, Coding and Analyzing Multimodality
- Session O31 - Multimodal Annotation
Chairperson : Jean Claude Martin
- Session P11 - Tools for Multimodal Corpus
Chair : Katerina Pastra
- Session P36 - Multimodal and Audiovisual Corpora
Chair : Daniel Sonntag

THE WORKSHOP THEMES

- **9:10 - 10:30 : HEAD MOVEMENTS / TOOLS (4 talks)**
- **11:00 - 13:00 : INTERACTION, EMOTION & CULTURE (6 talks)**
- **14:30 - 15:50 : MOTION CAPTURE AND VISION TECHNOLOGIES (4 talks)**
- **15:50 - 17:20 : POSTERS & DEMOS (10 posters & 4 demos)**
- **17:20 - 18:40 : GESTURE ANNOTATION & ANALYSIS (4 talks)**

- **Workshop Organisers**

Michael Kipp, DFKI, Germany

Jean-Claude Martin, LIMSI-CNRS, France

Patrizia Paggio, Univ. of Copenhagen, Denmark

Dirk Heylen, Univ. of Twente, The Netherlands

SOME MM TITLES

- Towards an Integrated Scheme for Semantic Annotation of Multimodal Dialogue Data
- Speech Grammars for Textual Entailment Patterns in Multimodal Question Answering
- Providing Multilingual, Multimodal Answers to Lexical Database Queries
- A Software Toolkit for Viewing Annotated Multimodal Data Interactively over the Web
- Work on Spoken (Multimodal) Language Corpora in South Africa
- Multimodal Russian Corpus (MURCO): First Steps
- The NOMCO Multimodal Nordic Resource - Goals and Characteristics
- The PlayMancer Database: A Multimodal Affect Database in Support of Research and Development Activities in Serious Game Environment
- BabyExp: Constructing a Huge Multimodal Resource to Acquire Commonsense Knowledge Like Children Do

THE D64 CORPUS



MASSIVE DATA RESOURCES

Multimodal Expressive data (d64)

◆◆◆audio◆◆◆◆◆

unless you really need the original recording, try the mp3 first - for easier downloads!

	day1 am		day1 pm		day2	
left field	(big)	mp3	(big)	mp3	(big)	mp3
right field	(big)	mp3	(big)	mp3	(big)	mp3
mix1	-	a b	-	x	-	a b c d
mix2	-	a b	-	y	-	a b c d
Nike	(big)	mp3	(big)	mp3	-	-
Catha	(big)	mp3	(big)	mp3	(big)	mp3
Jens	(big)	mp3	(big)	mp3	(big)	mp3
Fred	(big)	mp3	(big)	mp3	(big)	mp3
Nick	(big)	mp3	(big)	mp3	(big)	mp3
table	???	???	???	???	(big)	mp3
sofa	-	a b	-	y	-	a b c d
sony-b66	-	a b	-	y	-	a b c d
head-mount					(big)	nick
sync	(big)	-	(big)	-	(big)	-

these tables were filled from logic rather than listening ...
from wiring diagrams rather than ears-on experience
they are bound to be wrong!!

◆◆◆video◆◆◆◆◆

most of these are compressed to mp4 - should download MUCH faster than the originals

	day1 am		day1 pm		day2	
sofa	all	cut	-	cut	-	a ??
armchair	a b c	-	a b c d e	-	-	a b c d e f g h i j k
chair	uncut	-	-	-	-	clearing up
sofa left	all	-	-	-	uncut	-
fred	-	1 2	-	3 4	-	5 6 7 8

see note above .. same applies here!

◆◆◆extras◆◆◆◆◆

	day1 am	day1 pm	day2
loops		(fred)	
360-degree camera	a b c d e f	a b c d e f g h i j k l m n o p q r s t u v w x y z aa ab ac ad ae	a b c d e f g
other round	1 2	3a 3b	4 5
mocap	set 1	set 2	set 3
4-track bkp audio	?1 ?2	x1 x2 x3 x4 x5 x6 x7 x8 x9	xa xb xc xd xe

◆◆◆photos◆◆◆◆◆

see [here](#) (see separate page)

NEW & ARISING ISSUES

- alignment, storage & distribution of massive data
- crowd-sourcing of annotations
- podcasting & variable-size representations
- integration of modalities (gesture & speech / face & voice)
- naturalness of captured data (environments & ‘tasks’)

SUMMARY

- LREC MultiModal: 33 papers 5%
- annotation methodologies
- standards
- tools
- evaluation

MULTIMODALITY

- how could we have lived without it?
- watch this space!