Max, our Agent in the Virtual World

A Machine that Communicates with Humans

Ipke Wachsmuth

University of Bielefeld

University of Bielefeld

Collaborative Research Center SFB 360 and
Artificial Intelligence & Virtual Reality Lab

Situated Artificial Communicators
SFB 360

Technische Fakultät
Labor für Künstliche Intelligenz & Virtuelle Realität

sound check

Collaborative Research Center SFB 360

started in July 1993, overall funding by Deutsche Forschungsgemeinschaft
Directors: Prof. Gert Rickheit
Prof. Ipke Wachsmuth
`www.sfb360.uni-bielefeld.de`

SFB 360 Thematic fields

- Speech and Visual Perception
- Perception and Reference
- Knowledge and Inference
- Speech-Action Systems

Disciplines involved

- Linguistics
- Psycholinguistics
- Psychology
- Informatics
- Neuroinformatics
- Artificial Intelligence

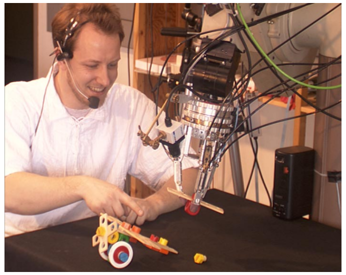
Leading research questions

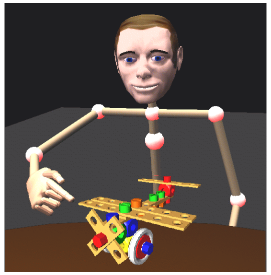
Situated Artificial Communicators SFB 360

How do humans communicate in a cooperative task robustly and successfully?

What can be learned from this about particular features of human intelligence?

Can we transfer communication abilities to artificial systems of robotics and virtual reality?



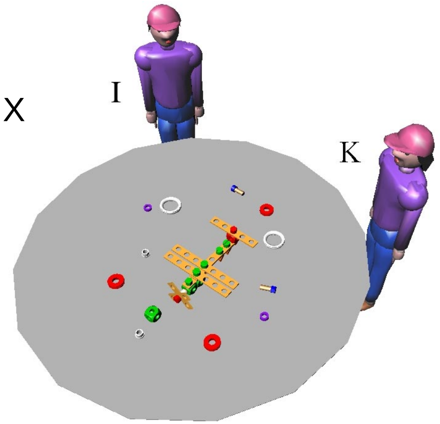


Scenario for investigation

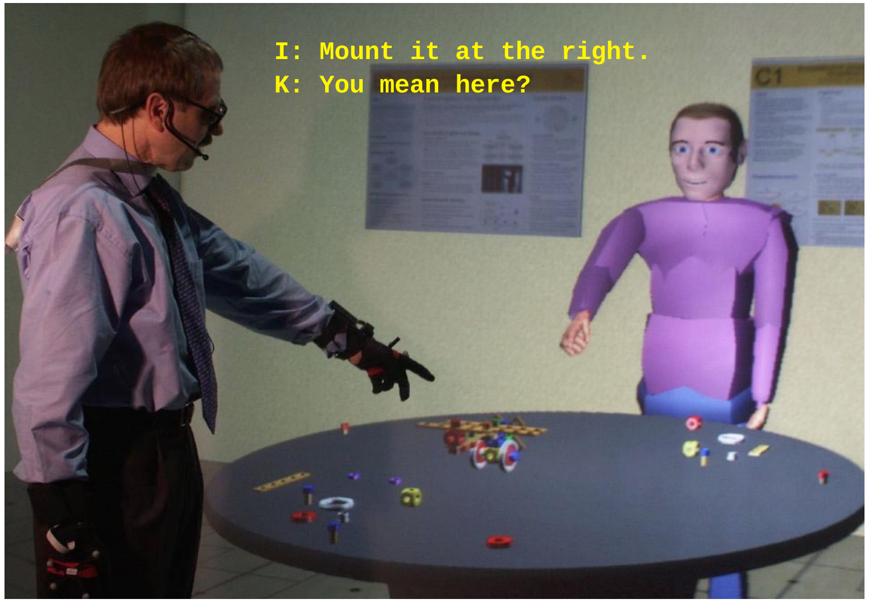
As most cognitive abilities are decisively situated, a specific reference situation serves to investigate task-oriented discourse.

For illustration the assembly of a model aeroplane from the BAUFIX construction kit is used.

A human instructor (I) and an artificial constructor (K) cooperate by way of an „Instructor-Constructor Dialog“



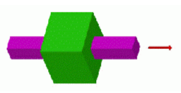
Situated communication



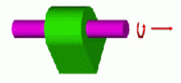
I: Mount it at the right.
K: You mean here?

Virtual Constructor

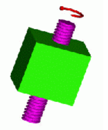
- „*lower kinematic pairs*“ can be modeled (uncoupled or coupled)
- based on Roth‘s 1994 book: „Konstruieren mit Konstruktionskatalogen“



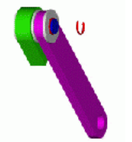
prismatic pair



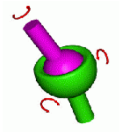
cylindrical pair



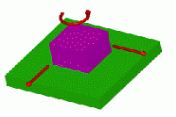
helical pair



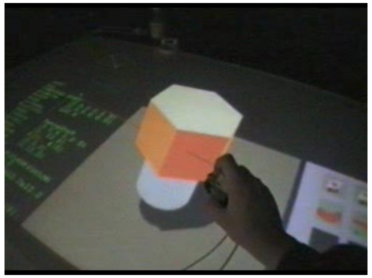
revolute pair



spherical pair



planar pair

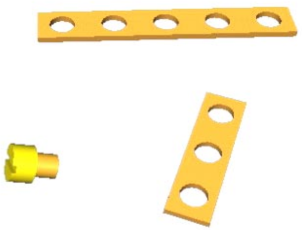


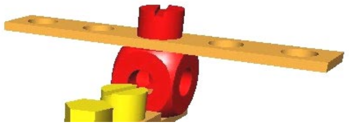
Virtual Constructor

Everything buildable with the 'BAUFIX' kit can be built in virtual reality.

Structural descriptions adapted dynamically:

- object descriptions are updated to comply with current situation
- make actual conceptualization available for dialogue





(e.g., 'bar' gets to be 'tail unit')

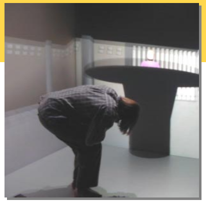
COAR representation formalism
Artificial Intelligence Review 10(3-4), 1996

Artificial Intelligence & Virtual Reality

Lab Research Mission

- AI methods used to establish an intuitive communication link between humans and multimedia
- Highly interactive Virtual Reality by way of multi-modal input and output systems (gesture, speech, gaze)
- Scientific enquiry and engineering of information systems closely interwoven (cognitive modeling approach)

New lab inaugurated 15 July ’02



- 3-sided Cave-like display
- 6 D-ILA projectors
- Passive stereo, circular-polarisation filters
- Gesture tracking: marker-based infrared-camera system
- precision hand posture tracking by two wireless datagloves
- 8-channel spatial sound system

… and Max

Applications

Embodied Conversational Agents [Cassell et al. 2000]

Computer-generated characters that demonstrate human-like properties in „face-to-face“ communication. Three aspects:

Multimodal Interfaces with natural modalities like speech, facial displays, hand gestures, and body stance

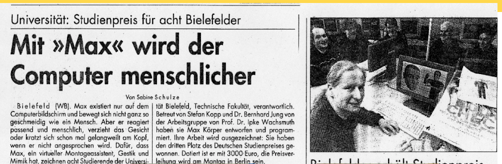
Software Agents that represent the computer in an interaction with a human or represent their human users in a digital environment (as „avatars“)

Dialog Systems where verbal as well as nonverbal devices advance the human-machine dialog

Technische Fakultät

Labor für Künstliche Intelligenz & Virtuelle Realität

„Brains“



Universität: Studienpreis für acht Bielefelder

Mit »Max« wird der Computer menschlicher

Von Sabine Schulze

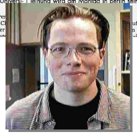
Bielefeld (WB). Max existiert nur auf dem Computerbildschirm und bewegt sich nicht ganz so geschmeidig wie ein Mensch. Aber er reagiert passend und menschlich, verzieht das Gesicht oder kratzt sich schon mal gelangweilt am Kopf, wenn er nicht angesprochen wird. Dafür, dass Max, ein virtueller Montageassistent, Gestik und Mimik hat, zeichnen acht Studierende der Universität Bielefeld, Technische Fakultät, verantwortlich. Betreut von Stefan Kopp und Dr. Bernhard Jung von der Arbeitsgruppe von Prof. Dr. Ipke Wachsmuth haben sie Max Körper entworfen und programmiert. Ihre Arbeit wird ausgezeichnet: Sie haben den dritten Platz des Deutschen Studienpreises gewonnen. Dotiert ist er mit 3000 Euro, die Preisverleihung wird am Montag in Berlin sein.

Das Thema des Forschungswettbewerbs für Studierende lautet »Bodycheck – Wieviel Körper braucht der Mensch?« eingereicht

B… Studienpreis

Ipke Wachsmuth

Bernhard Jung


Marc Latoschik

Timo Sowa

Ian Voß

Peter Biermann

Stefan Kopp

Alf Kranstedt

Nadine Leßmann

Computers

- ‘Artabel Fleye 160’ Linux Cluster for application and rendering
- 5 server nodes (double-Pentium III-class PCs)
- 8 graphic nodes (single-Pentium IV-class PCs) with NVIDIA GeForce 3 graphics
- nodes linked via 2GBit/s Myrinet-network for distributed OpenGL rendering

– thanks for €s to DFG –

Agent MAX

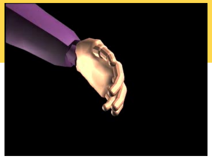
Situated Artificial Communicators
SFB 360

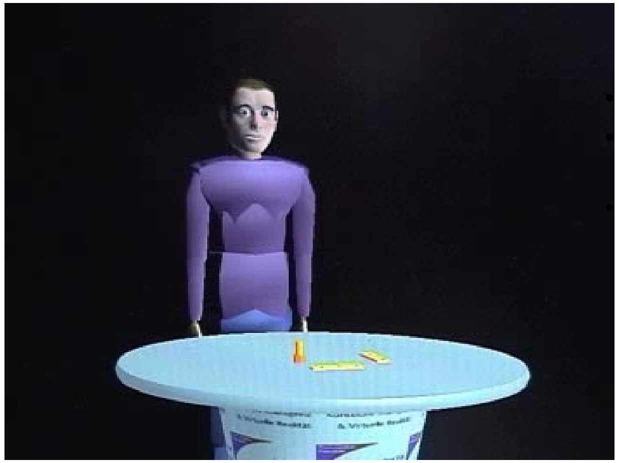
An artificial communicator situated in virtual reality

Research into fundamentals of communicative intelligence:

- PHYSIS – the body system (especially gestures)
- COGNITION – the knowledge system
- EMOTION – the valuation system

PHYSIS: Articulated body



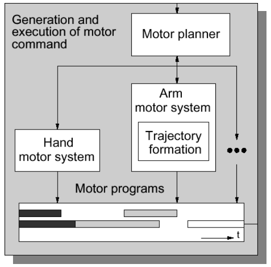


Kinematic skeleton with 53 degrees of freedom (DOF) in 25 joints for the body and 25 DOF for each hand

Hand animated by key framing

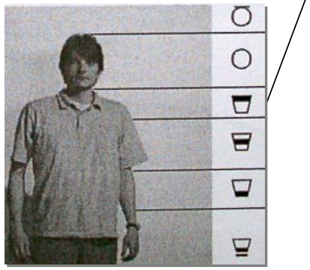
Body animated by model-based animation

Motion generators running concurrently and synchronized



HamNoSys for gesture form description

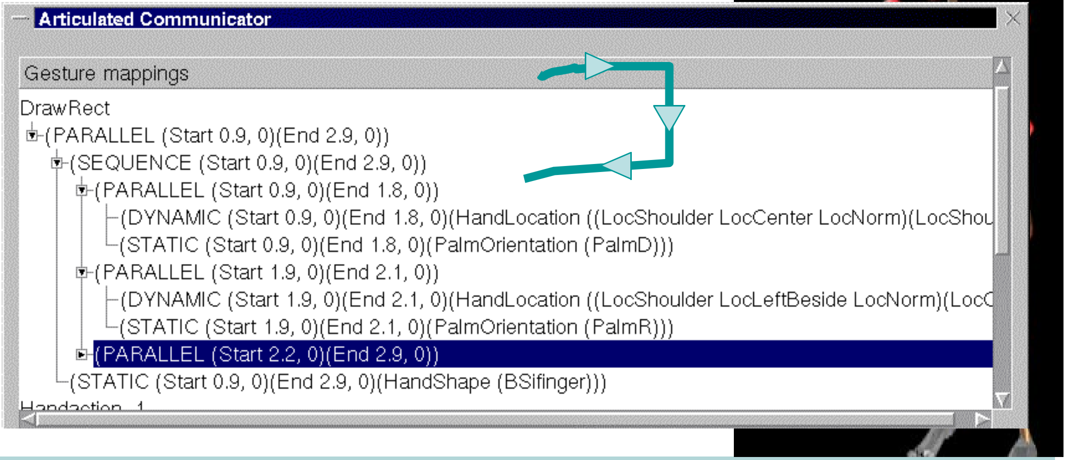
(„Hamburg Notation System“ – Institut für Deutsche Gebärdensprache, Hamburg)



Symbol	ASCII-equivalent	Description
⅃	BSifinger	indexfinger stretched
⏏	EFinA	extended ahead
0	PalmL	palm orientated left
⊔	LocShoulder	location shoulder height
↳»	LocStretched	fully stretched out
↥	MoveA	hand move ahead
→	MoveR	hand move right
<etc.>	...	...
( )	( )	executed in parallel
[ ]	[ ]	executed in sequence

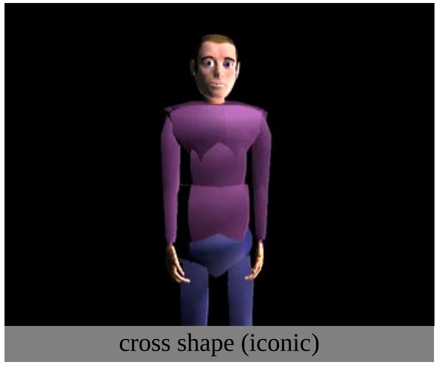
Outlining a (roughly) rectangular shape

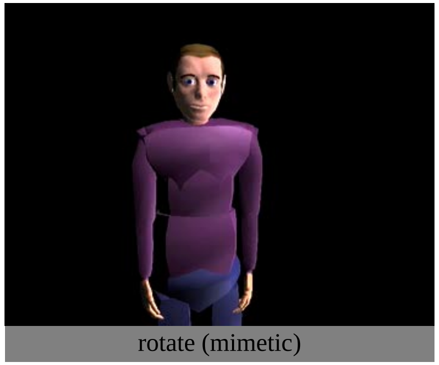
HamNoSys + movement constraints + timing constraints
(selected) {STATIC, DYNAMIC} {Start, End, Manner}

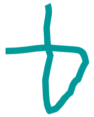


```
DrawRect
⊟(PARALLEL (Start 0.9, 0)(End 2.9, 0))
  ⊟(SEQUENCE (Start 0.9, 0)(End 2.9, 0))
    ⊟(PARALLEL (Start 0.9, 0)(End 1.8, 0))
      ├(DYNAMIC (Start 0.9, 0)(End 1.8, 0)(HandLocation ((LocShoulder LocCenter LocNorm)(LocShou
      └(STATIC (Start 0.9, 0)(End 1.8, 0)(PalmOrientation (PalmD)))
    ⊟(PARALLEL (Start 1.9, 0)(End 2.1, 0))
      ├(DYNAMIC (Start 1.9, 0)(End 2.1, 0)(HandLocation ((LocShoulder LocLeftBeside LocNorm)(LocC
      └(STATIC (Start 1.9, 0)(End 2.1, 0)(PalmOrientation (PalmR)))
    ⊞(PARALLEL (Start 2.2, 0)(End 2.9, 0))
  └(STATIC (Start 0.9, 0)(End 2.9, 0)(HandShape (BSifinger)))
Handaction_1
```

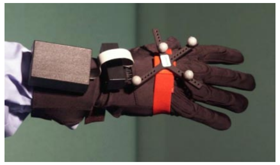
More form gestures


cross shape (iconic)


rotate (mimetic)

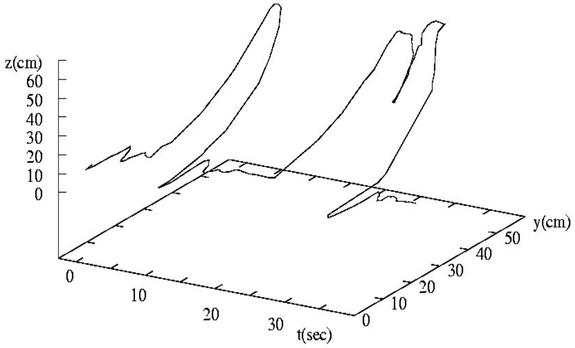


Measuring gestures

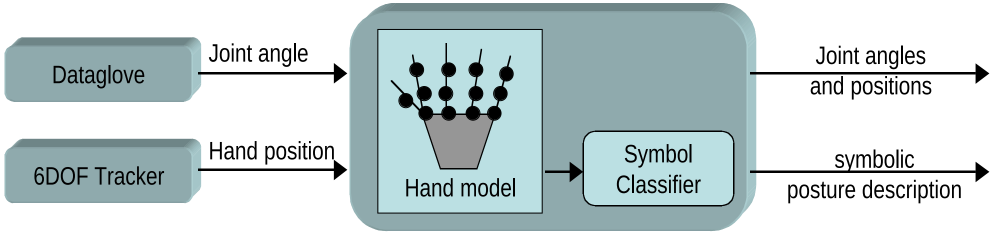


Segmentation cues

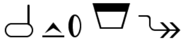
- strong acceleration of hands, stopps, rapid changes in movement direction
- strong hand tension
- symmetries in two-hand gestures



Analyzing gestures

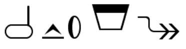


Symbolic classification of gesture *shape* (HamNoSys)



Gesture imitation game

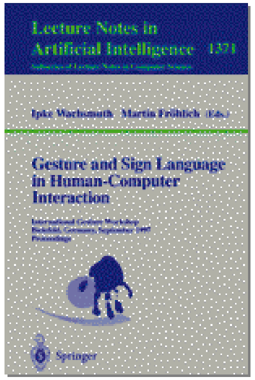
- Human displays gestures, Max imitates them
- Parsing of gesture input: HamNoSys
- HamNoSys for specification of gesture output



Real time!

2 Gesture books (1998, 2002)

LNAI 1371



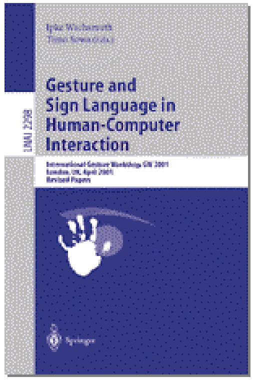
Lecture Notes in Artificial Intelligence 1371

Ipke Wachsmuth Martin Fröhlich (Eds.)

Gesture and Sign Language in Human-Computer Interaction

Springer

LNAI 2298

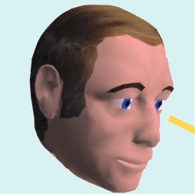


Gesture and Sign Language in Human-Computer Interaction

LNAI 2298

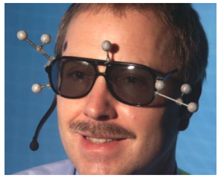
Springer

COGNITION: Analyzing language



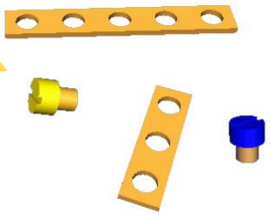
I: Steck die gelbe Schraube in die lange Leiste.

- **speech recognition**
- **syntactic-semantic parsing**
- **reference to perceived scene**



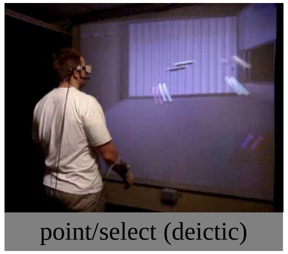
Insert the yellow bolt into the long bar.

steck	COMMAND	CONNECT
die	DET	
gelbe	COLOR	YELLOW
Schraube	OBJECTTYPE	BOLT
in	PREP	IN
die	DET	
lange	SIZE	LARGE
Leiste	OBJECTTYPE	BAR

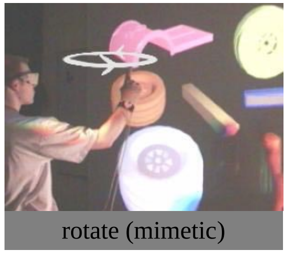


```
(select x (OBJECTTYPE(x)= BOLT and COLOR(x)= YELLOW))
(select y (OBJECTTYPE(y)= BAR  and SIZE(y) = LARGE))
```

Multimodal Analysis: tATN



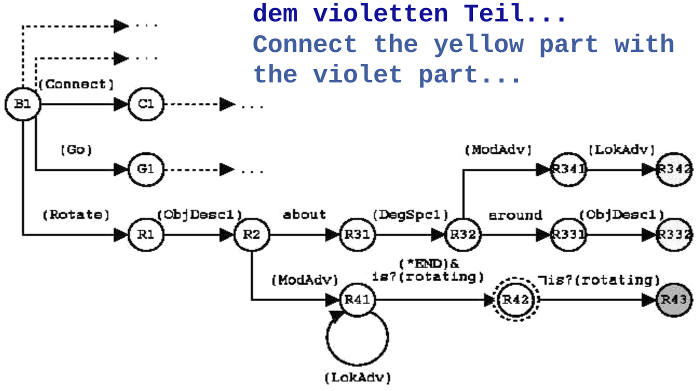
point/select (deictic)



rotate (mimetic)

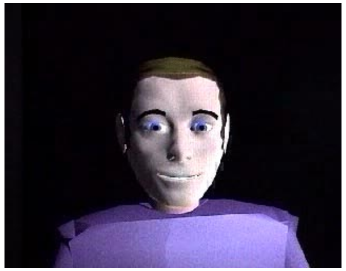
Verbinde das gelbe Teil mit dem violetten Teil...

Connect the yellow part with the violet part...



- Integration of speech and gesture
- Interpretation in application context

Lip-synchronous speech



Text-to-Speech:

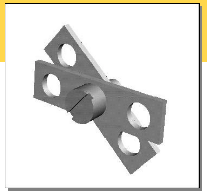
TXT2PHO (IKP Uni Bonn), MBROLA Phoneme transcription is the basis for automatic generation of visemes.

(Concept-to-Speech: TO DO)

- one viseme for *M, P, B*
- one viseme for *N, L, T, D*
- one viseme for *F, V*
- one viseme for *K, G*
- plus visemes for the vowels

Historical: Zemanek-Vocoder

Speech and accentuation



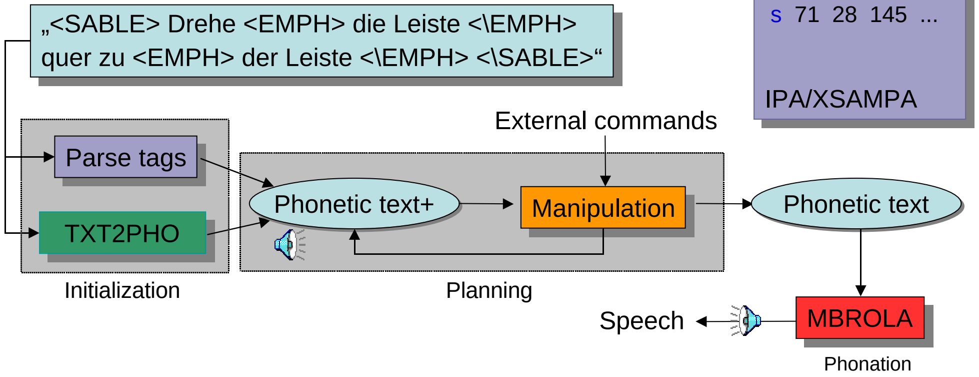
Drehe die Leiste quer zu der Leiste.
Turn this bar crosswise to that bar.

Drehe *die* Leiste quer zu *der* Leiste.
Turn *this* bar crosswise to *that* bar.

Phonetic text:

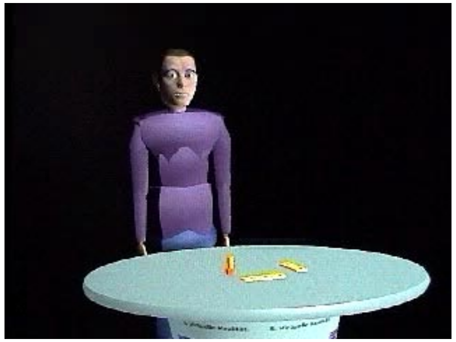
s	105	18	176	...
p	90	8	153	
a:	104	4	150	...
s	71	28	145	...

IPA/XSAMPA



„<SABLE> Drehe <EMPH> die Leiste <\EMPH> quer zu <EMPH> der Leiste <\EMPH> <\SABLE>“

Uttering speech and gesture



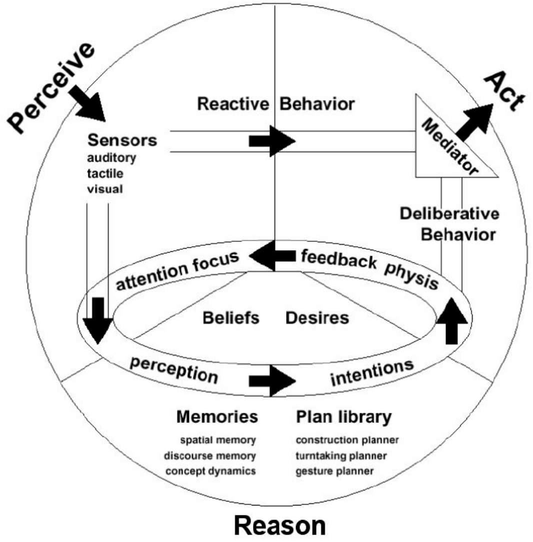
And now take this bar and make it this big.

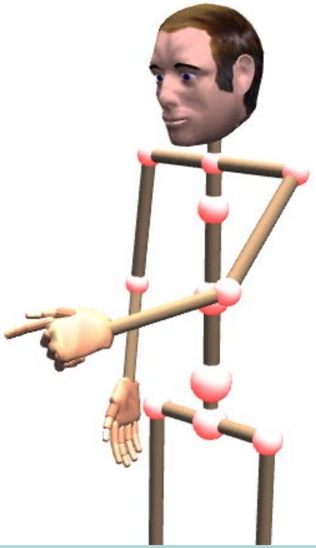
MURML: XML-based markup language for multimodal utterance representations

```
<utterance>
  <specification>
    Und jetzt nimm <time id="t1"/> diese Leiste
    <time id="t2“ chunkborder="true"/>
    und mach sie <time id="t3“/> so gross. <time id="t4"/>
  </specification>
  <behaviorspec id="gesture_1">
    <gesture>
      <affiliate onset="t1" end="t2"/>
      <constraints>
        <parallel>
          <static slot="HandShape" value="BSifinger"/>
          <static slot="ExtFingerOrientation"
               value="$object_loc_1“ mode="pointTo"/>
          <static slot="GazeDirection" value="$object_loc_1"
               mode="pointTo"/>
        </parallel>
```

Cognitively motivated architecture

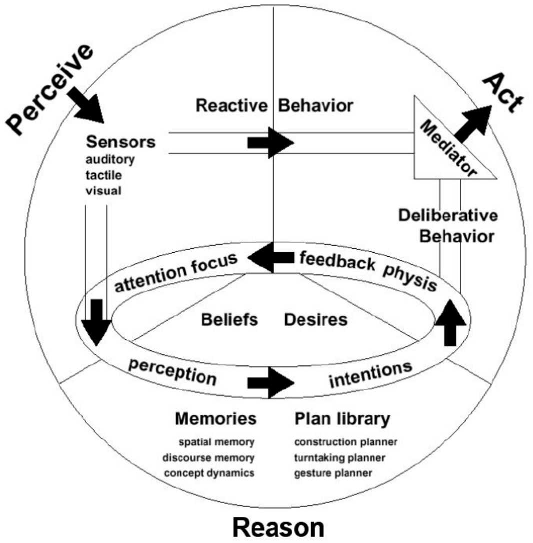
Situated Artificial Communicators
SFB 360





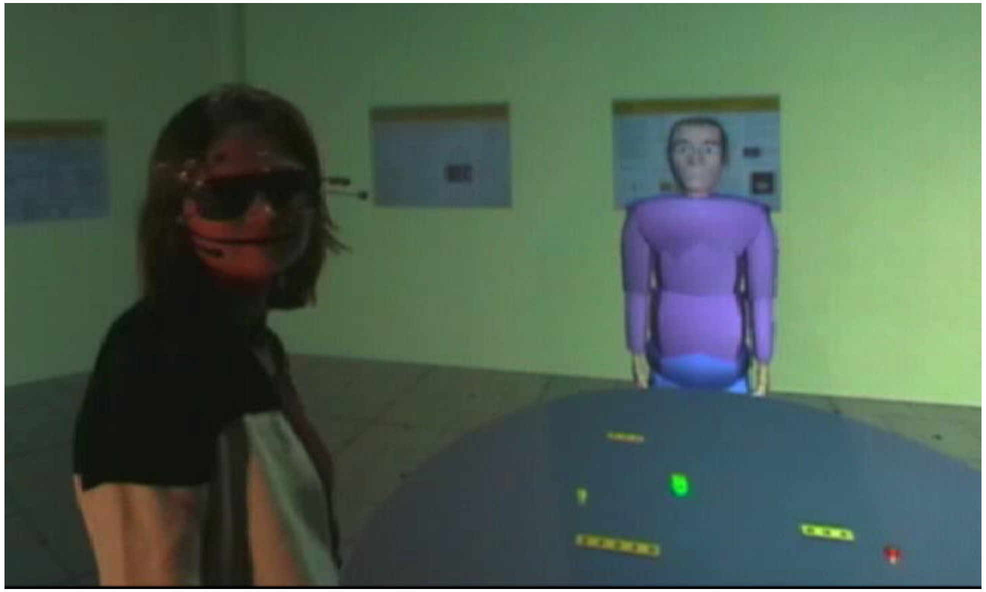
Cognitively motivated architecture

Situated Artificial Communicators SFB 360

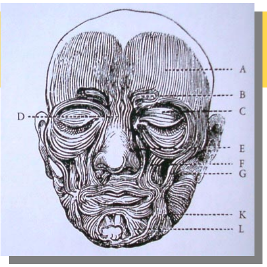


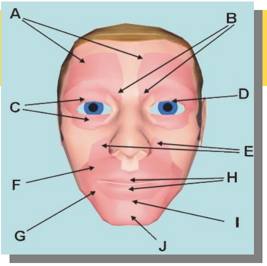
- Perceive, Reason, Act running concurrently
- parallel processing by a reactive and a deliberative system
- information feedback in a cognitive loop
- BDI kernel with selfcontained dynamic planners
- account for embodiment (physis) of the agent, multimodality

Communicating with Agent Max...



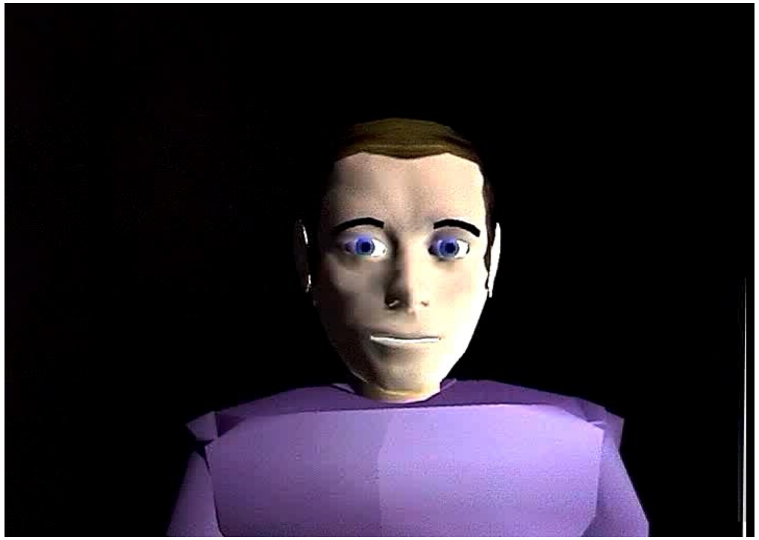
Facial expression





Muskeln des linken Bildes (von Sir Ch. Bell)		Virtuelle Muskeln des rechten Bildes (MAX)	
A	Stirnmuskel	A	Stirnmuskel
B	Augenbrauenrunzler	B	Augenbrauenrunzler
C	Augenringmuskel	C	Augenringmuskel
D	Pyramidenmuskel der Nase	D	Augenlidmuskel
E	Heber der Oberlippe u d. Nasenflügels	E	Heber der Oberlippe u d. Nasenflügels
F	eigentlicher Lippenheber	F	Jochbeinmuskel u. Mundwinkelheber
G	Jochbeinmuskel	G	Mundwinkelherabzieher
K	Mundwinkelherabzieher	H	Ringmuskel des Mundes
L	Viereckiger Kinnmuskel	I	Unterlippenherabzieher
		J	Unterkiefer

Expression of EMOTION

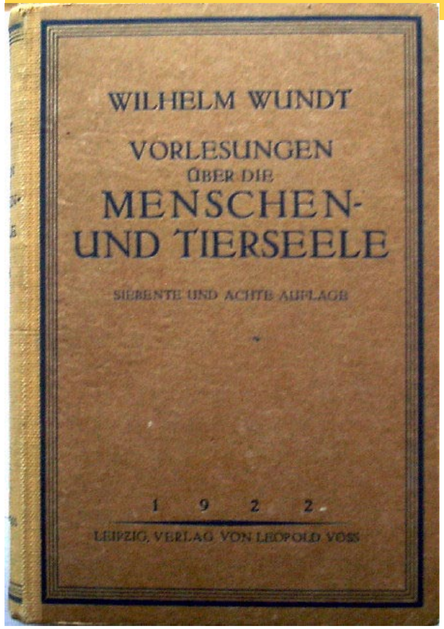


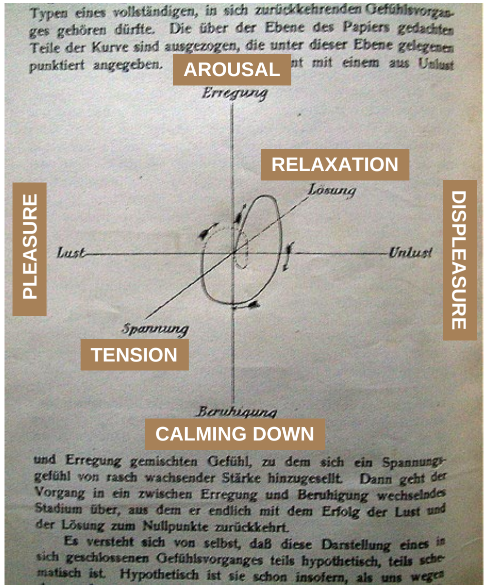
Coordinated control of face muscles based on Action Units (Ekman/Friesen)

Student project (Körber Prize!)

Emotive system under development

Wundt Emotion Dynamics



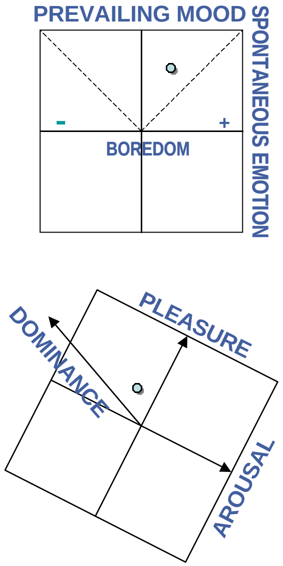

Typen eines vollständigen, in sich zurückkehrenden Gefühlsvorganges gehören dürfte. Die über der Ebene des Papiers gedachten Teile der Kurve sind ausgezogen, die unter dieser Ebene gelegenen punktiert angegeben. …nt mit einem aus Unlust

AROUSAL
Erregung
RELAXATION
Lösung
PLEASURE
Lust
Unlust
DISPLEASURE
Spannung
TENSION
Beruhigung
CALMING DOWN

und Erregung gemischten Gefühl, zu dem sich ein Spannungsgefühl von rasch wachsender Stärke hinzugesellt. Dann geht der Vorgang in ein zwischen Erregung und Beruhigung wechselndes Stadium über, aus dem er endlich mit dem Erfolg der Lust und der Lösung zum Nullpunkte zurückkehrt.

Es versteht sich von selbst, daß diese Darstellung eines in sich geschlossenen Gefühlsvorganges teils hypothetisch, teils schematisch ist. Hypothetisch ist sie schon insofern, als uns wegen

Affect – Dynamic Emotion Space



Embodied Communication

Anthropomorphic appearance

- humanoid body
- personality
- facial expression
- gesture
- spoken language
- emotional features

Intentionality

- knowledge / beliefs
- desires / motivations
- intentions
- commitments
- emotions...

e.g., BDI architecture ++
(Beliefs - Desires - Intentions)

THE END