

Multimodal Interaction in Virtual Reality

Ipke Wachsmuth
Faculty of Technology
University of Bielefeld

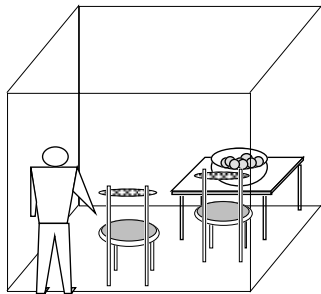


Technische Fakultät
 Labor für Künstliche Intelligenz & Virtuelle Realität
 SG IM Speech and Gesture Interfaces for Multimedia
 Virtuelle Wissensfabrik
 SFB 360 Sited Artificial Communicators
 NRW
 Lecture Notes in Artificial Intelligence 1371
 Ipke Wachsmuth, Martin Freilich (Eds.)
 Gesture and Sign Language in Human-Computer Interaction
 International Gesture Workshop, Bielefeld, Germany, September 1997
 Springer

This contribution reports work carried out in Bielefeld in the context of interacting with virtual reality environments.

Anthropomorphic Interfaces

- Interfaces following human form
- natural human-machine communication
- highly popular in interactive multimedia
- gestures, speech, coverbal gesture



Put this chair there!

Communicating in the situation

Central aspects in our work:

- Multimedia
- Multimodality
- Multiple representations
- Multi-agent Systems

Drawbacks in VR

Virtual Reality:

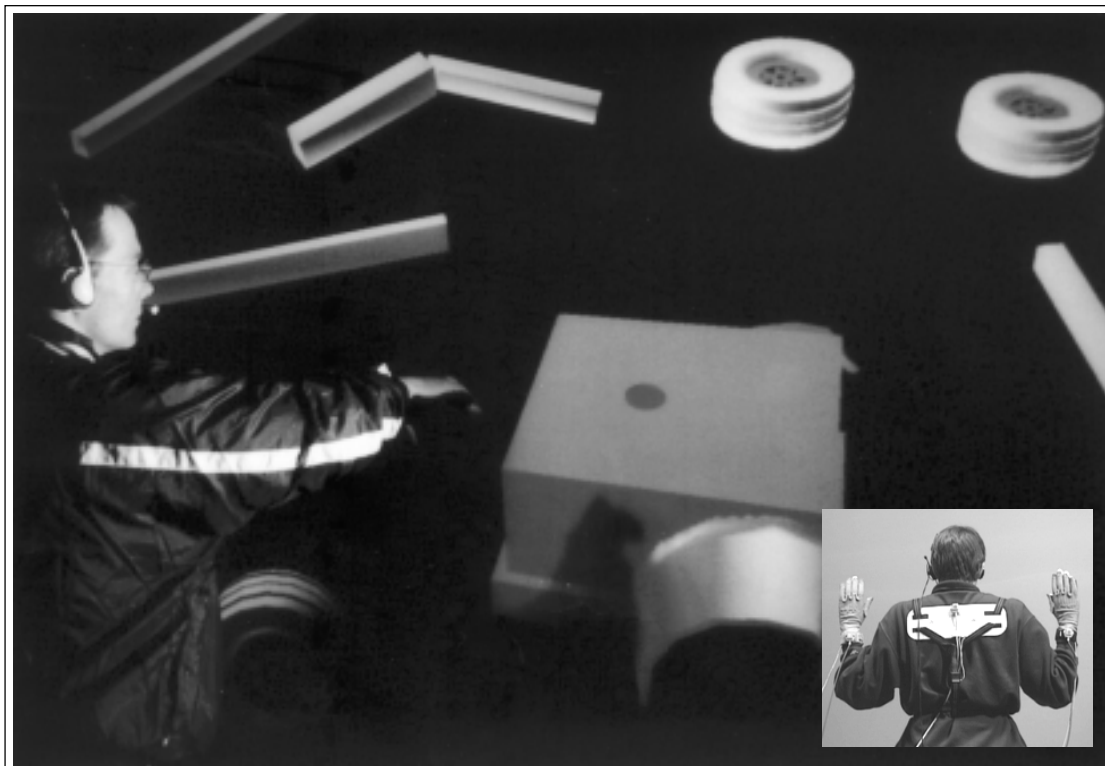
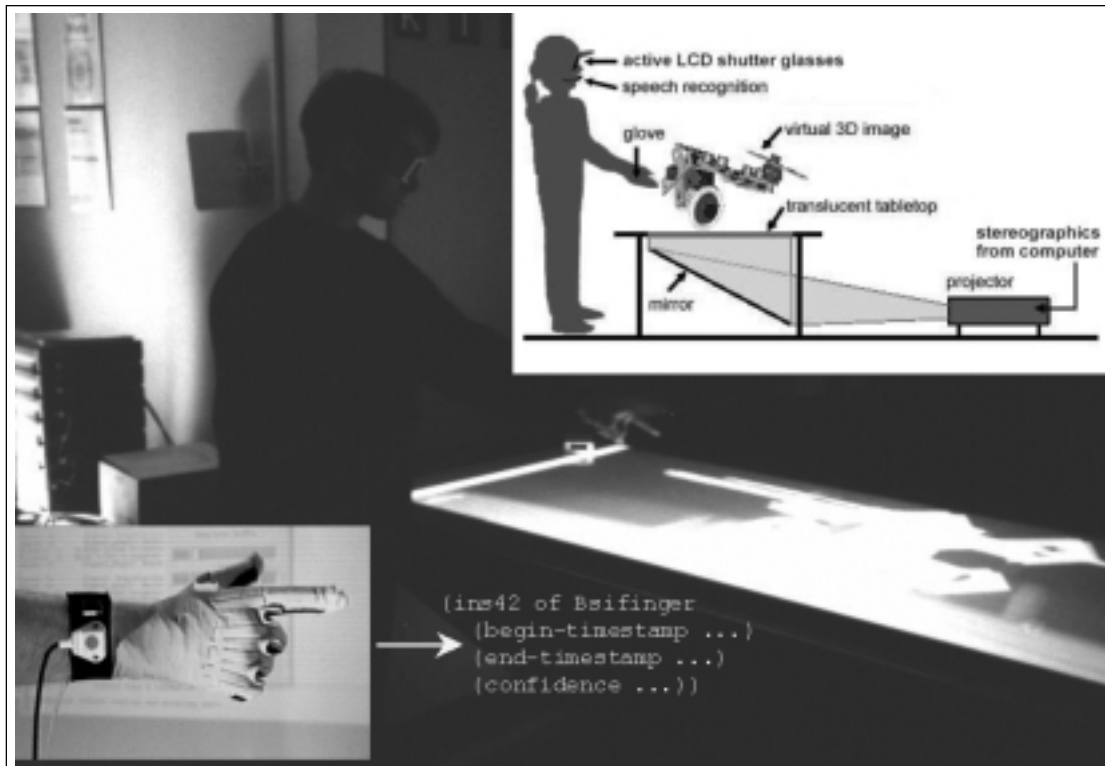
- 3D viewing
- spatial sound
- tactile stimuli

- ◆ Tedious usage of synthetic models
- ◆ Deficits in interactive design support
- ◆ Manipulation not subject to physical laws
- ◆ Lack of user-friendliness in interaction

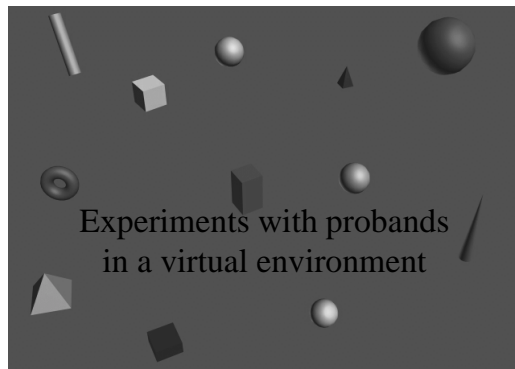
Our approach: indirect management techniques, based on gesture and speech

Three things are important in our work toward incorporating gestures as a useful tool in virtual reality:

- measuring gestures as articulated hand and body movements in the context of speech
- interpreting them by way of classifying features and transducing them to an application command via a symbolic notation inherited from sign language
- timing gestures in the context of speech in order to establish correspondence between accented behaviors in both speech and gesture channels



Work with
Timo Sowa
and Ian Voss



Gesture Recognition

Approach 1: Basic interactions, direct feedback

- ◆ Pointing
- ◆ Grasp
- ◆ GraspRelease
- ◆ Rotation
- ◆ Translation

Approach 2: Formbased description

- ★ explicit (knowledge-based) approach
- ★ atomic form elements of gesture are composed to 'gesture words'

Symbols for Body Parts

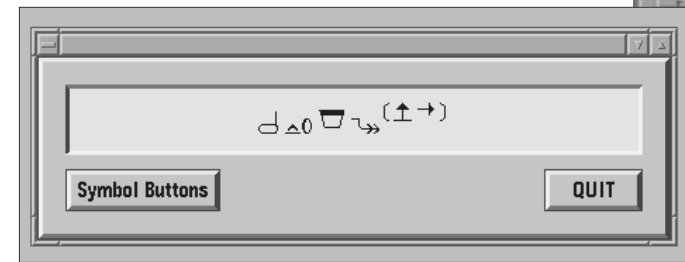
- 5 4 3 2 1 \
- finger tip
- finger nail
- finger joint
- knuckle joint
- base of finger
- side of finger
- ball of thumb
- back of hand
- side of small finger
- palm
- side of thumb



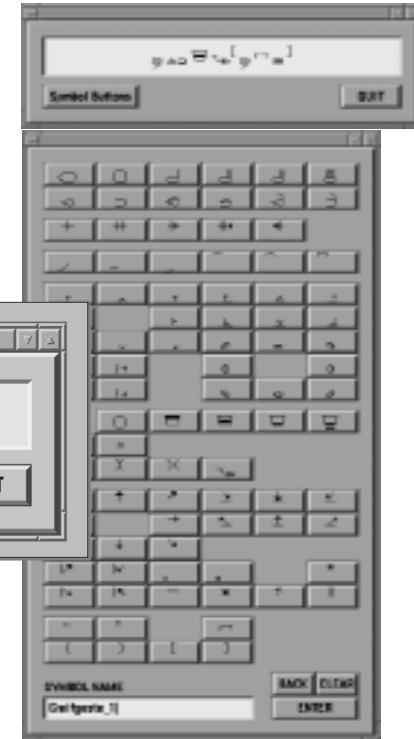
- head
- top of head
- forehead
- ~ eyebrows
- eyes
- ↑ nose
- ∇ ear
- }) cheek
- ← mouth
- ∪ chin
- ∪ under the chin
- ∪ neck
- ∪ shoulder
- ∪ upper arm
- ∪ chest
- ∪ outside elbow
- ∪ inside elbow
- ∪ lower arm
- back of wrist
- pulse
- ∪ stomach
- ∪ below stomach

Gesture notation with HamNoSys

Work with Martin Fröhlich and
Timo Sowa, SGIM Project



For form description of gestures we work with a scenario-specific subset of HamNoSys (HNS').



```
(defrule rule32
?tmp28 <- (object (is-a MoveA)) ; eine Handvorwärtsbewegung
?tmp31 <- (object (is-a MoveR)) ; eine Handrechtsbewegung
=>
(SEQUENCE tmpc125 ?tmp28 ?tmp31)) ; in Sequenz
```

Hier ist automatisch eine CLIPS-Regel rule32 generiert worden, die ein Token der Klasse tmpc125 assertieren kann, das eine erkannte Sequenz von bestimmten Handbewegungen registriert.

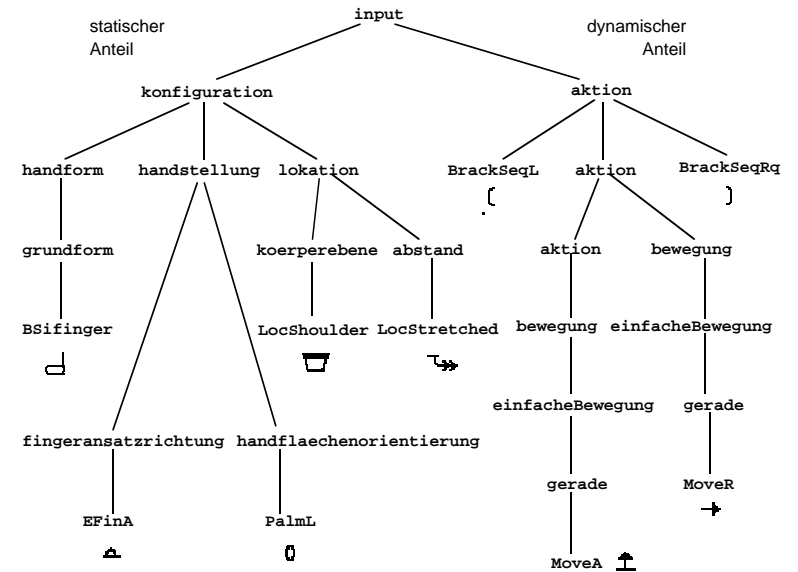


```
(defrule rule36
?tmp16 <- (object (is-a BSifinger)) ; Zeigefinger gestreckt
?tmp19 <- (object (is-a EFinA)) ; nach vorn gestreckt
?tmp22 <- (object (is-a PalmL)) ; Handfläche links
?tmp23 <- (object (is-a LocShoulder)) ; in Schulterhöhe
?tmp24 <- (object (is-a LocStretched)) ; ausgestreckt
?tmp35 <- (object (is-a tmpc125)) ; bewegt wie oben
=>
(PARALLEL DUMMY ?tmp16 ?tmp19 ?tmp22 ?tmp23 ?tmp24 ?tmp35))
```

Hier ist automatisch eine Regel rule36 generiert worden, die ein Token der Klasse DUMMY assertieren kann, das das Auftreten einer bestimmten Handpostur in Verbindung mit oben beschriebener Bewegung registriert.

(für DUMMY kann ein beschreibender Gestenname gesetzt werden)

HNS' Parse Tree for



From gesture to application

- ◆ **Actuators**
 - abstract placeholders of significant discrete features in body reference system; normalized (e.g., world coordinates)
- ◆ **Motion modifiers**
 - bind temporarily to actuators and filter motion data to object transformations (via manipulators)
- ◆ **Manipulators**
 - receive transformation commands and put them into effect in the 3D scene

Multimodal Interface:

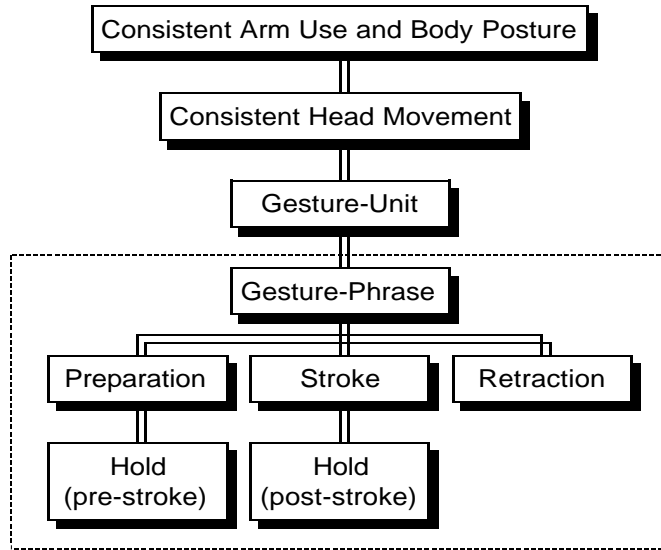
an interface which allows the(sequential or concurrent) usage of multiple modalities

Open input:

Begin and end of interaction not known

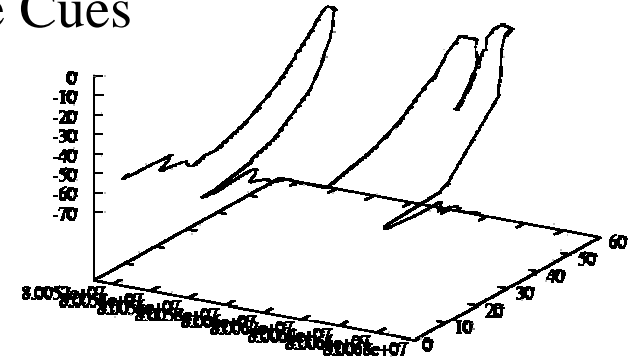
Gestures: Kinesic Structure

After McNeill, Levy & Pedelty (1990)



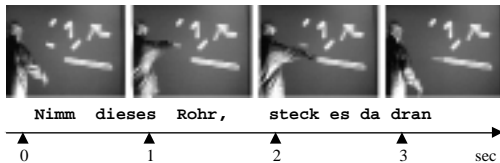
Stroke Cues

- pre/post-stroke hold (kinesic structure)
- strong acceleration of hands, stops, rapid changes in movement direction
- strong hand tension
- symmetries in two-hand gestures



Timing of gestures and speech

- The gesture stroke is often marked by an abrupt stop which is correlated with accented words or syllables
- the stroke does not occur *after* an accented word but simultaneously or shortly before



=> hypotheses for establishing correspondence between accented behaviors in speech and gesture channels

Multimodal Integration

Two "logistic" problems to be solved (Srihari, 1995):

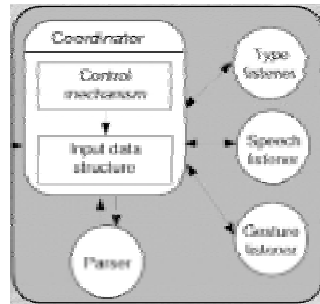
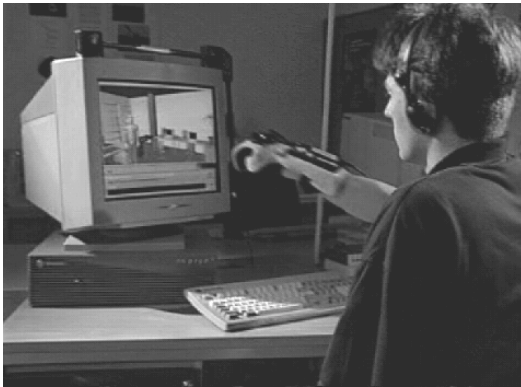
► Segmentation Problem

How can a system be made to cope with 'open input'? How can units be determined to be processed in one system cycle?

► Correspondence Problem

How to determine cross-references between multiple modalities (speech/gesture)?

Earlier work: VIENA (1996)

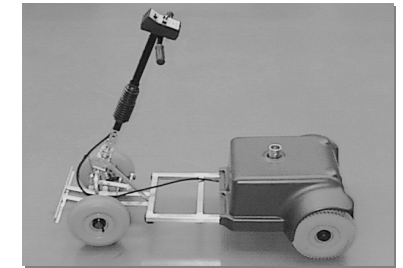
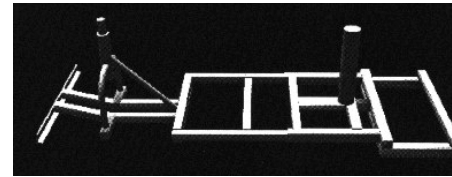


put | this | computer | on | the | blue | desk

Virtual Construction

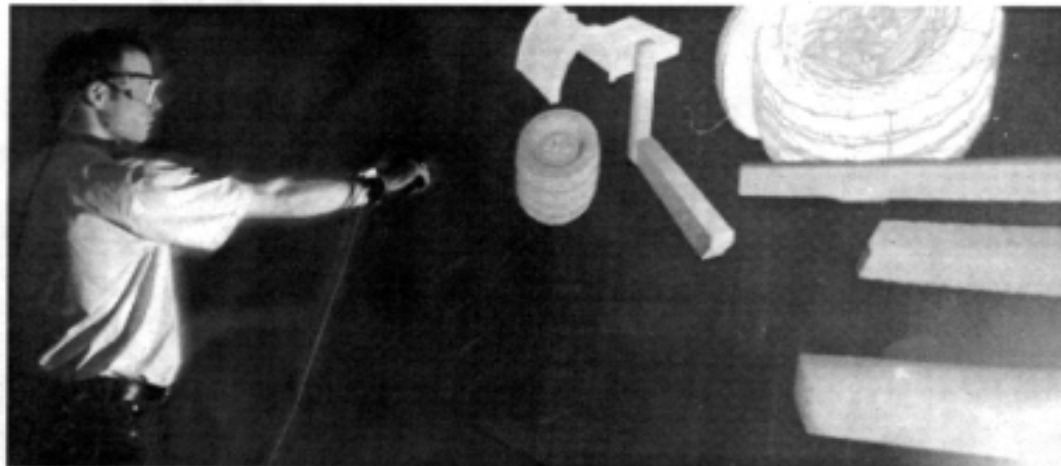
Work with Bernhard Jung and Martin Hoffhenke

- Innovative approach to prototype the design of mechanical devices and explore their assembly in early stages of industrial development, based on CAD models



The things we have learned from investigating these issues help us to advance natural interaction with 3D stereographic scenes in a scenario of virtual construction.

11. 3. 99 StadtBlatt 11



Wie von Geisterhand: Stefan Kopp bewegt Projektionen auf der Leinwand per Handstreich

Foto: Konrad Lehmann

Das Weltentor in der Uni

Informatiker bauen an der Virtuellen Realität

Eines der literarischen Ereignisse 1998 war «Otherland» von Tad Williams, gefeiert als der Roman der Virtuellen Realität: Eingepackt in Datenslots oder versenkt in gefüllte Säuge, bewegen sich die Hei-

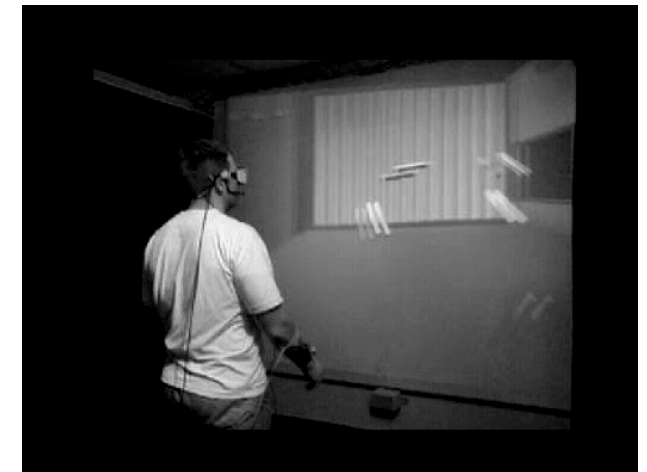
formatiker sich nicht nur mit Physik und Ingenieurwesen beschäftigen, sondern auch mit Psychologie, Biologie, Soziologie, Linguistik.

Eines der literarischen Ereignisse 1953 war «Fahrenheit 451»

Multimodal Interaction in VR

Work with Marc Latoschik, SGIM Project

SGIM Speech and
Gesture
Interfaces
for Multimedia



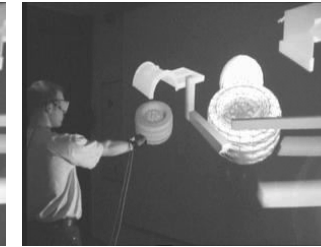
Natural interaction

Work with Marc Latoschik, SGIM Project



pointing/selecting
(deictic)

turning
(mimetic)



drag — and — drop

In the first place we have dealt with pointing and turning, etc., commonly classified as deictic and mimetic gestures.



Feuertaufe: Ohne Computer-Absturz ging die Präsentation von Marc Latoschik (Foto rechts) über die Bühne. Mit 3D-Brillen auf der Nase verfolgten die Gratulanten die Bilder im neuen Labor. FOTO: BILDREISE

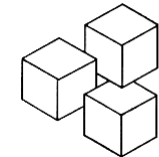
Where we would like to go next



- ◆ Virtual Workspace (two-sided 3D projection)
- ◆ manipulative gesture (grasp space) and
- ◆ communicative gesture (distant space)

DEIKON (2000++)

Deixis in Construction dialogs
(joint project with Hannes Rieser)



Situated
Artificial
Communicators

SFB 360

- ◆ Systematic study of referential acts by coverbal gesture
- ◆ Aim: To investigate how complex signals originate from speech and gesture and how they are used in reference
- ◆ Contribution of gestural deixis for making salient or selecting objects and regions
- ◆ Making an artificial communicator to understand and produce (coverbal) deictic gestures in construction dialogs

Deictics & Iconics

Work with Timo Sowa and Marc Latoschik



In the DEIKON project, we have now started to research into more sophisticated forms of deictics in construction dialogues that include features indicating shape or orientation, which leads us into iconic gesture.

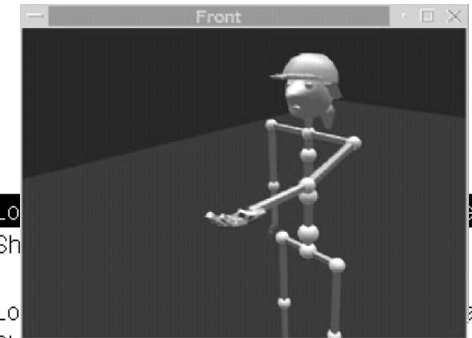
Lifelike Gesture Synthesis

Work with Stefan Kopp,
Articulated Communicator

Pull-1

```

    (PARALLEL (Start 1.1, 0)(End 2.3, 0))
    (SEQUENCE (Start 1.1, 0)(End 2.3, 0))
    (PARALLEL (Start 1.1, 0)(End 1.7, 0))
    (STATIC (Start 1.1, 0)(End 1.7, 0)(HandLo
    (STATIC (Start 1.1, 0)(End 1.7, 0)(HandSh
    (PARALLEL (Start 1.9, 0)(End 2.3, 0))
    (STATIC (Start 1.9, 0)(End 2.3, 0)(HandLo
    (STATIC (Start 1.9, 0)(End 2.3, 0)(HandShape BSist))
    (STATIC (Start 1.1, 0)(End 2.3, 0)(PalmOrientation PalmD))
    
```

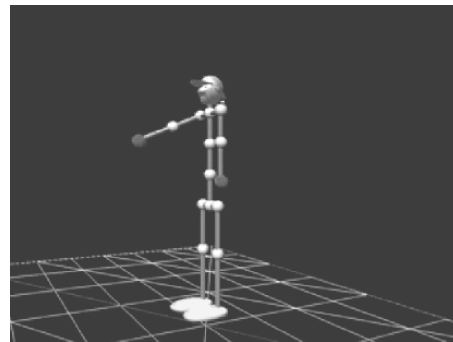


Another issue in our work is the synthesis of lifelike gesture from symbolic descriptions for an articulated virtual figure where natural motion and timing are central aspects.

Find Lab Showcase & papers...

Technische
Fakultät

Labor für
Künstliche Intelligenz
& Virtuelle Realität



www.techfak.uni-bielefeld.de/techfak/ags/wbski/
www.techfak.uni-bielefeld.de/~ipke/

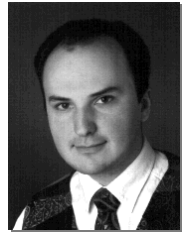
WBS Team



Bernhard Jung



...Hoffhenke



...Fröhlich

Martin...



Stefan Kopp



Timo Sowa



Ian Voß



Marc Latoschik