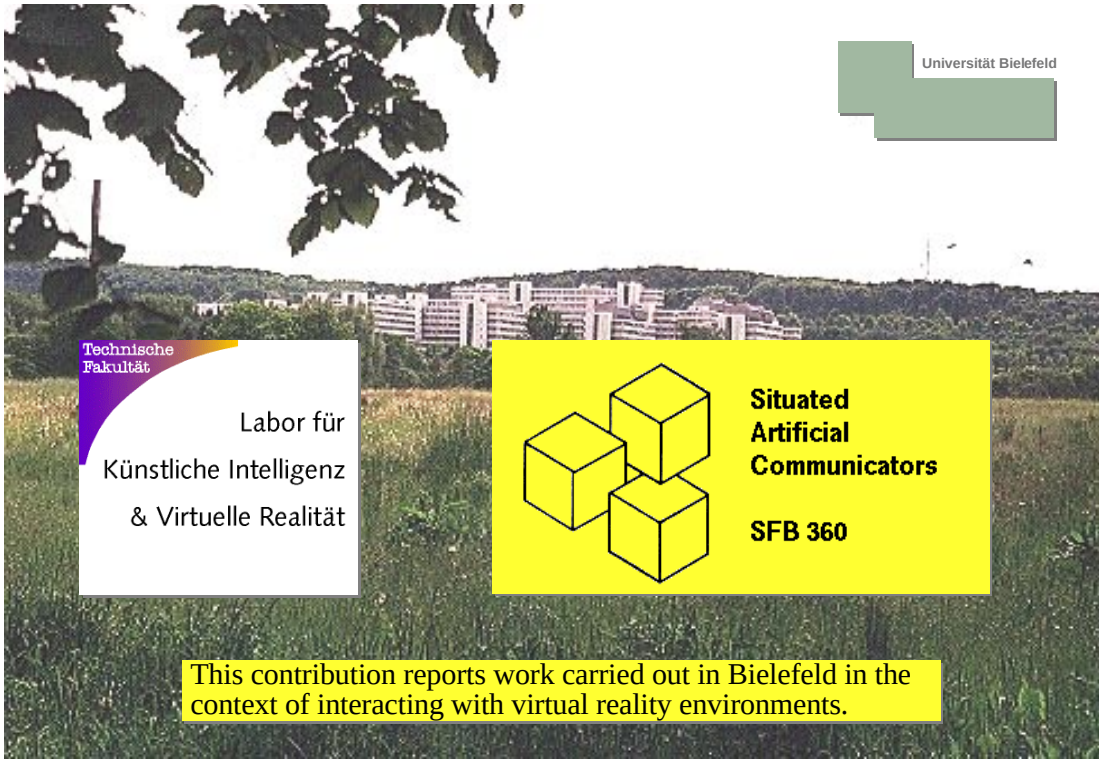
Universität Bielefeld

Lifelike Gesture Synthesis and Timing for Conversational Agents

Ipke Wachsmuth

Faculty of Technology
University of Bielefeld



Universität Bielefeld

Overview

- **Previous Work and Context**
- **Lifelike Gestures Synthesis**
- **Articulated Communicator**
- **Timing**
- **Outlook: Speech and Facial Gesture**

sound check

Universität Bielefeld

Natural Interaction in VR

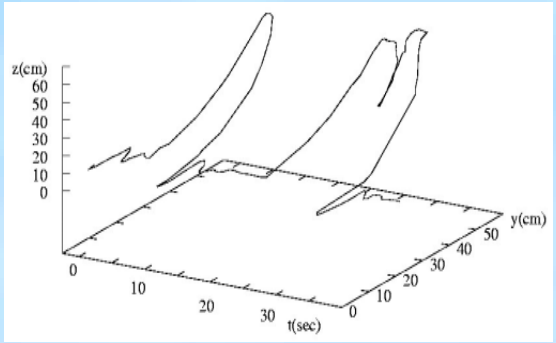
Three things have been important in our work toward incorporating gestures as a useful tool in virtual reality:

- 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

Universität Bielefeld

Measuring gestures; stroke cues

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



Universität Bielefeld

From gesture to application

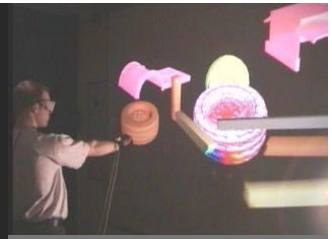
```
(ins42 of Bsifinger
    (begin-timestamp ...)
    (end-timestamp ...)
      (confidence ...))
```

- **Actuators**
  - abstract placeholders of significant discrete features in body reference system; normalized in world coordinates
- **Motion modificators**
  - 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

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.

Universität Bielefeld

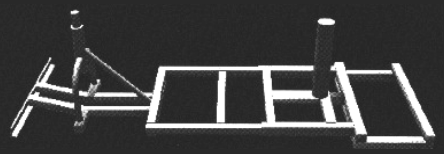
Virtual Laboratory



pointing/selecting (deictic)

turning (mimetic)

Speech and Gesture in Virtual Construction
Work with Marc Latoschik & Bernhard Jung



Universität Bielefeld

Deictics & Iconics

Work with Timo Sowa, Marc Latoschik and Ian Voss

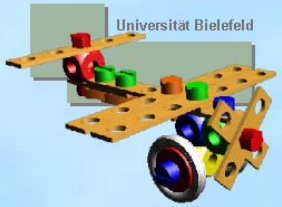
Situated Artificial Communicators

SFB 360

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 lead into iconic gesture.

Example: Outlining a rectangular shape

Universität Bielefeld



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 able to understand and produce (coverbal) deictic gestures in construction dialogs

Universität Bielefeld

Overview

- **Previous Work and Context**
- **Lifelike Gestures Synthesis**
- **Articulated Communicator**
- **Timing**
- **Outlook: Speech and Facial Gesture**

Universität Bielefeld

Lifelike gesture synthesis - why?

RATIONALES

„Learning to generate is learning to understand“

- better understanding of biologic and of cognitive factors of communication abilities through a generative approach
- Models of explanation in the form of „biomimetic“ simulations

New type of advanced application interfaces

- Synthesis of lifelike gesture is finding growing attention in HCI
- Generation of synthetic gesture in connection with text-to-speech systems is one of the goals for embodied conversational agents

Embodied Conversational Agents => new paradigm for the study of gesture and for human-computer interface!

Universität Bielefeld

Embodied Conversational Agents

(after Cassell, Sullivan, Prevost & Churchill 2000)

- Computer-generated characters that demonstrate similar properties as humans in „face-to-face“ conversation, incl. abilities for verbal and nonverbal communication
- ECAs can be viewed as
  - multimodal interface with natural modalities like speech, facial displays, hand gestures, and body stance
  - software agents, insofar as they represent the computer in an interaction with a human or represent their human users in a computational environment (as „avatars“)
  - dialogue system, where verbal as well as nonverbal devices advance the human-computer dialogue

Universität Bielefeld

Focus of this talk

Synthesis of lifelike gestures for an articulated virtual agent, with particular emphasis on how to achieve temporal coordination with external information such as the signal generated by a text-to-speech system.

Early work; ongoing!

Universität Bielefeld

Gestures in Humans

(findings from various sources)

Gesture production

- abstract gesture specifications generated from spatiotemporal representations of „shape“
- transformed into patterns of control signals interpreted by low-level motor systems
- resulting gesture exhibits characteristical shape and dynamical properties

Gestural movements

- composed of distinct movement phases which form a hierarchical kinesic structure
- In coverbal gestures the stroke is closely coupled to accompanying speech
- semantic, pragmatic, and temporal synchrony between modalities

Universität Bielefeld

Research Challenges

How to generate communicative behaviors automatically?

- Skeletal animation for lifelike gesture synthesis
- Verbal behavior, also known as speech
- Facial animation and lip synch speech

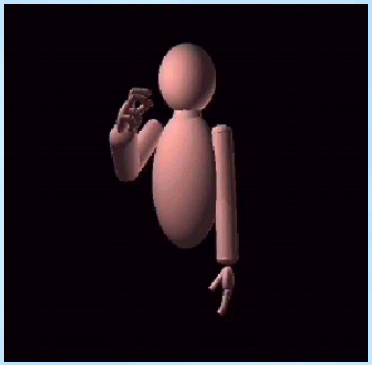
TIMING: achieving synchrony in accented behaviors, e.g., gesture stroke with stressed syllable

Universität Bielefeld

Example GeSSyCa System

(Lebourque & Gibet, Proceedings GW ’99)

- French sign language gestures produced from explicit movement representations
- based on a set of movement primitives (pointing, straight line, curved, circle, wave)
- combined to more complex gestures, reproduces natural movement characteristics
- adaptation of kinematic properties and timing as required in coverbal gesture not a focus



Virtual Signer (LIMSI-CNRS)

Universität Bielefeld

Example REA System

(„Real Estate Agent“; Cassell et al., 2000)

- Embodied conversational agent producing natural verbal and nonverbal outputs
- gesture animation schedules behaviors that, once started, cause several motor primitives to be executed
- based on keyframe animation and inverse kinematics
- exact timing of spoken and gestural utterances targetted, but not satisfactorily solved

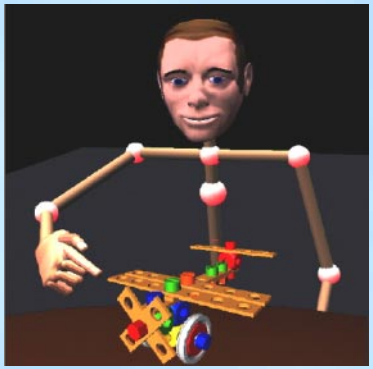
Agent REA (MIT Media Lab)

Universität Bielefeld

Example MAX System

(„Articulated Communicator“; Work with Stefan Kopp)

- Goal: real-time, lifelike gesture animation from representations of spatio-temporal gesture knowledge
- means of motion representation, planning, and control to produce multiple kinds of gestures
- gestures parametrized with re-spect to kinematics, i.e. velocity profile and overall duration, as well as to shape properties
- in addition: „peak timing“



Agent MAX (Bielefeld Lab)

Universität Bielefeld

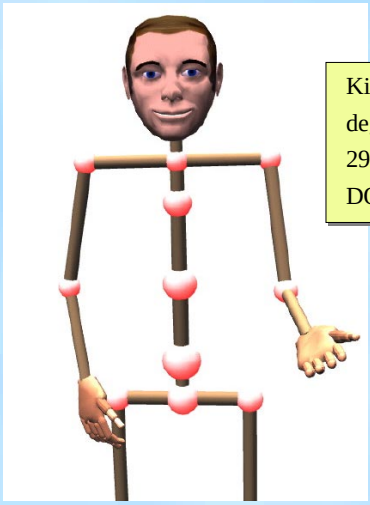
Overview

- **Previous Work and Context**
- **Lifelike Gestures Synthesis**
- **Articulated Communicator**
- **Timing**
- **Outlook: Speech and Facial Gesture**

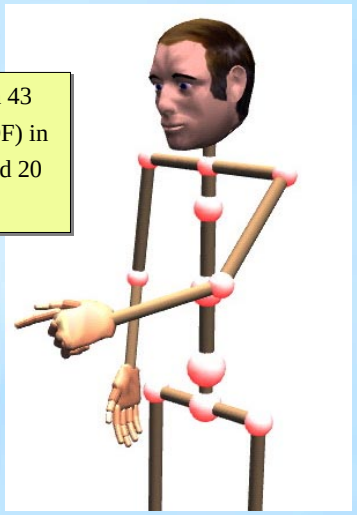
Universität Bielefeld

Articulated Communicator

Work with Stefan Kopp



Kinematic skeleton with 43 degrees of freedom (DOF) in 29 joints for the body and 20 DOF for each hand.



knees and feet not shown here

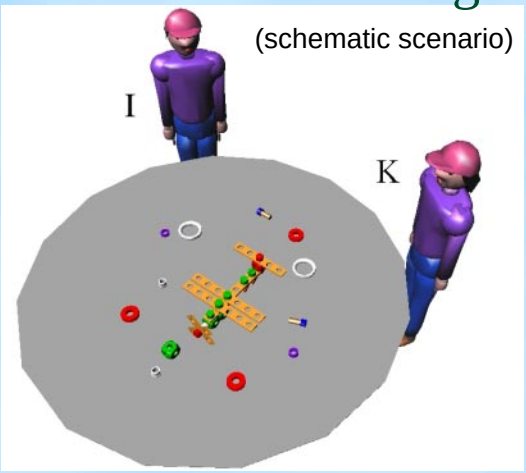
Universität Bielefeld

Instructor-Constructor Dialogue

Situated Artificial Communicators

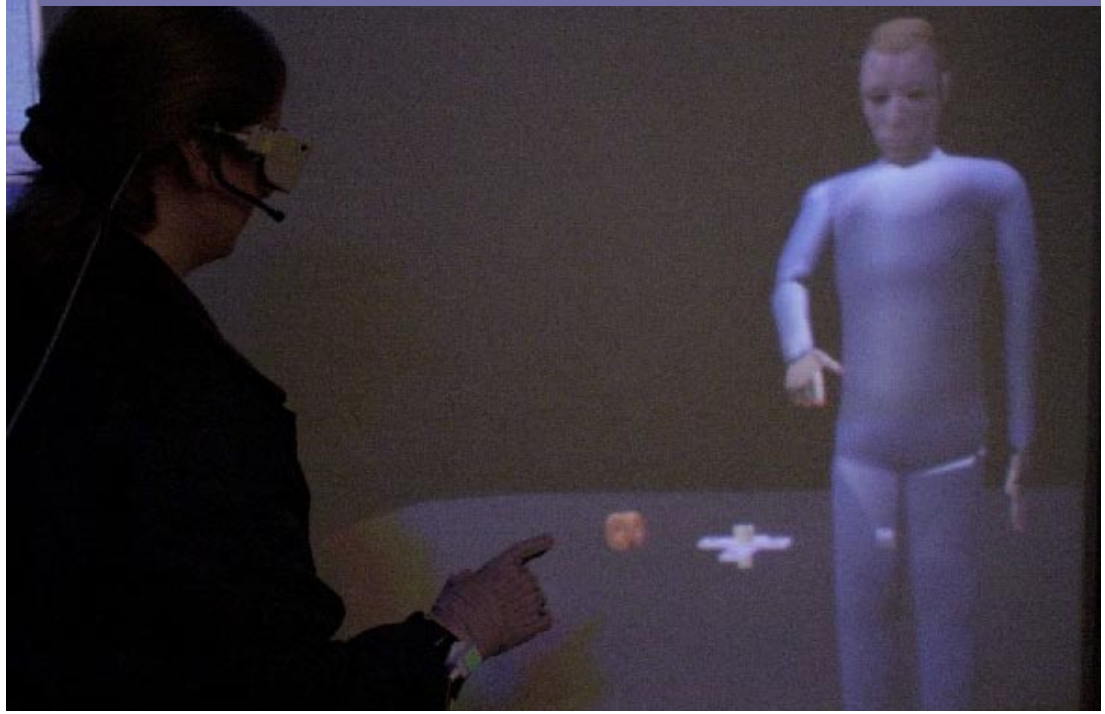
SFB 360

Scenario: Two communicators cooperate in constructing a model aeroplane



Two settings: Human instructor (I)
- constructor (K) is a robotic agent
- constructor (K) is a virtual agent

Articulated Communicator (target scenario)



Universität Bielefeld

Gesture animation

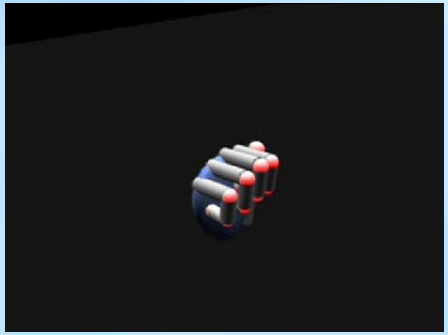
„Gestures ... exist because they have some distinctiveness in their Effort and Shape parameter.“ (Costa et al., 2000)

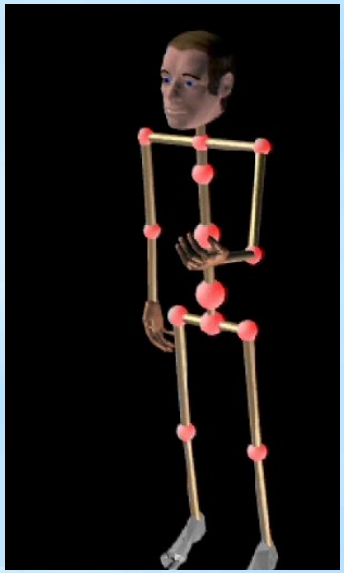
- **Flexibility**, **accuracy**, and **naturalness**!
- Two approaches to skeleton motion control:
  - Motion drawn from a database of predefined motions
  - **Motion calculated on demand**
- Integration of several motion generators vital for designing complex motions!
  - hand vs. arm movement
  - gesture stroke vs. retraction
  - emblematic vs. iconic gestures

Universität Bielefeld

Kinematic hand & body model

- Hand animated by key-framing
- Body animated by model-based animation
- motion generators running concurrently and synchronized





Universität Bielefeld

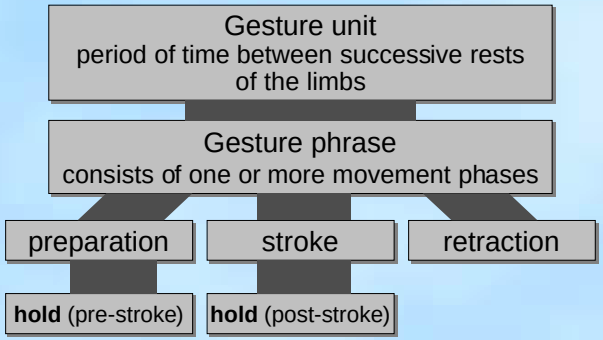
Overview

- **Previous Work and Context**
- **Lifelike Gestures Synthesis**
- **Articulated Communicator**
- **Timing**
- **Outlook: Speech and Facial Gesture**

Universität Bielefeld

Timing of gestures and speech

- Coverbal gestures are closely related to speech flow (semantic, pragmatic, and temporal synchrony; McNeill, 1992)
- Characteristic spatio-temporal features and kinematic properties



- Gestures co-occur with rheme (e.g. Cassell, 2000)
- Stroke onset precedes or co-occurs with the most *contrastively* stressed syllable in speech and covaries with it in time. (De Ruiter, 1998; McNeill, 1992; Kendon, 1986)

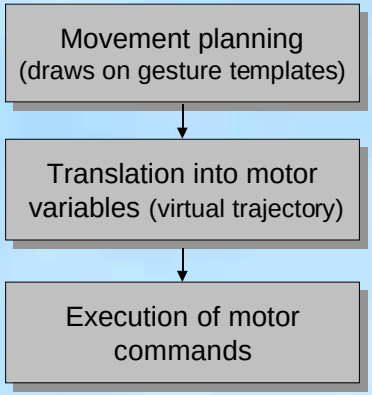
Universität Bielefeld

Gesture production & planning

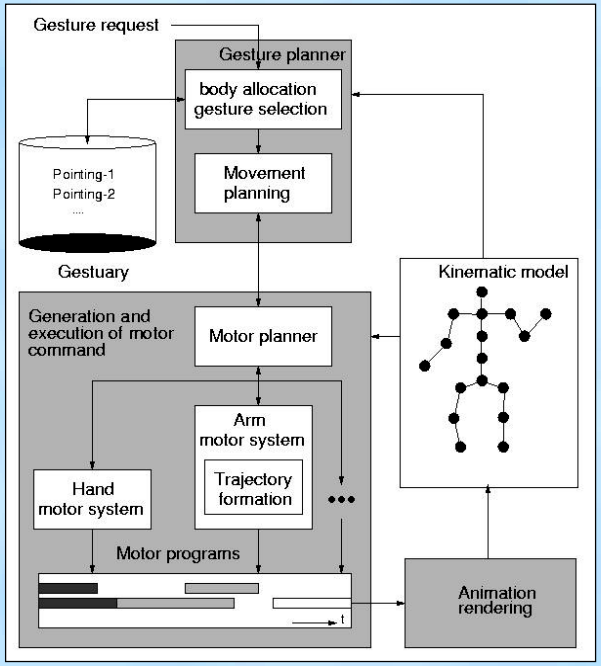
- Start from high-level, parametrizeable gesture representations
  - Script-based animations, e.g., PaT-Nets (Badler et al. 1993)
  - Feature-based descriptions based on some gesture/ movement notation system (Calvert et al., 1982; Lebourque & Gibet, 1999; Kopp & Wachsmuth, 2000)
- Planning of (voluntary) movements is performed
  - in external coordinates
  - in terms of significant trajectory properties
  - based on knowledge about initial and target locations
- Multistep planning process, feature-based description

Universität Bielefeld

Gesture animation



Build on HamNoSys for movement plans!



Universität Bielefeld

From templates to gesture plans

Movement planning
(draws on gesture templates)

Example gesture template from Gestuary

```
(IDENT    Pointing-2)
(FUNCTION    Refer_To_Loc (RefLoc RefDir))
(CONSTRAINTS
        (PARALLEL
            (STATIC (HandShape  (BSifinger)))
            (STATIC (HandLocation  (RefLoc)))
            (STATIC (ExtFingerOrientation (RefDir)))
            (STATIC (PalmOrientation (PalmD)))
        )
)
```

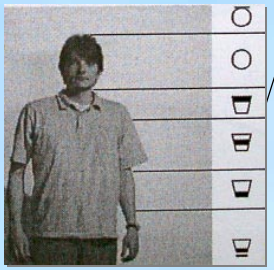
A movement plan is formed as a tree representation of a temporally ordered set of movement constraints in three steps:

- retrieve a feature-based specification from gestuary
- adapt it to the individual gesture context
- qualify temporal movement constraints in accordance with external timing constraints

Universität Bielefeld

Some HamNoSys Symbols

Work with Timo Sowa, based on Prillwitz et al. (1989) "Hamburg Notation System"



Symbol	ASCII notation	Description
	BSifinger	basis shape index finger stretched
	EFinA	extended finger orientation ahead
	PalmL	palm orientation left
	LocShoulder	location shoulder height
	LocStretched	location stretched
	MoveA	move hand ahead
	MoveR	move hand right
<etc.>	...	...
( )	PARALLEL	executed in parallel
[ ]	SEQUENCE	executed in sequence

Universität Bielefeld

Gesture plan notation (stroke)

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

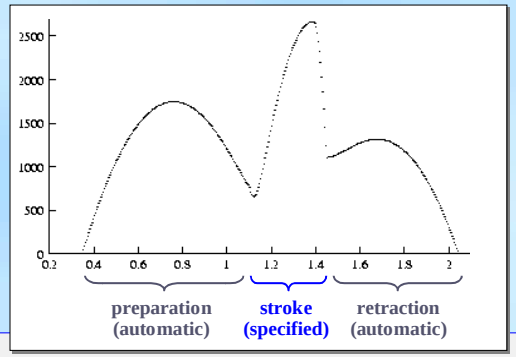
Gesture Specification = spatio-temporal feature-based description of mandatory parts of a gesture, i.e., the stroke

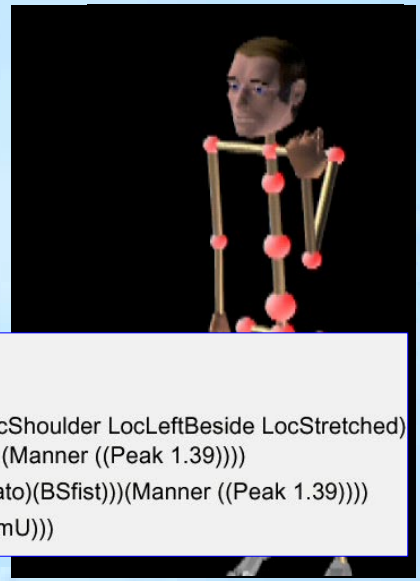
Example: „pull gesture“ stroke

```
Pull-2
└─(PARALLEL (Start 1.1)(End 1.41))
    ├─(DYNAMIC (Start 1.1)(End 1.41)(HandLocation ((LocShoulder LocLeftBeside LocStretched)
    │                (LocShoulder LocLeftBeside LocNear))) (Manner ((Peak 1.39))))
    ├─(DYNAMIC (Start 1.1)(End 1.41)(HandShape ((BSflato)(BSfist)))(Manner ((Peak 1.39))))
    └─(STATIC (Start 1.1)(End 1.41)(PalmOrientation (PalmU)))
```

Universität Bielefeld

Gesture plan notation (example)

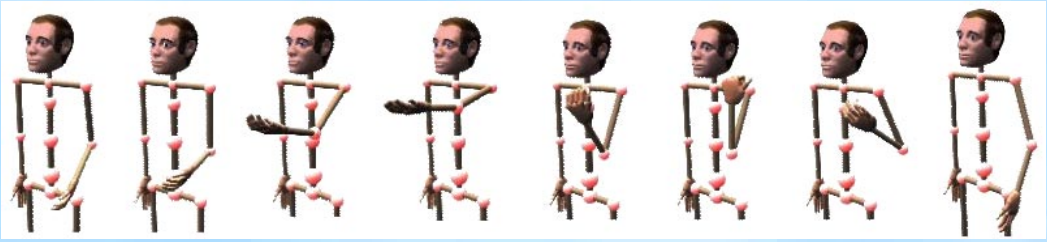




```
Pull-2
└─(PARALLEL (Start 1.1)(End 1.41))
    ├─(DYNAMIC (Start 1.1)(End 1.41)(HandLocation ((LocShoulder LocLeftBeside LocStretched)
    │                (LocShoulder LocLeftBeside LocNear))) (Manner ((Peak 1.39))))
    ├─(DYNAMIC (Start 1.1)(End 1.41)(HandShape ((BSflato)(BSfist)))(Manner ((Peak 1.39))))
    └─(STATIC (Start 1.1)(End 1.41)(PalmOrientation (PalmU)))
```

Universität Bielefeld

Gesture plan execution (phases)



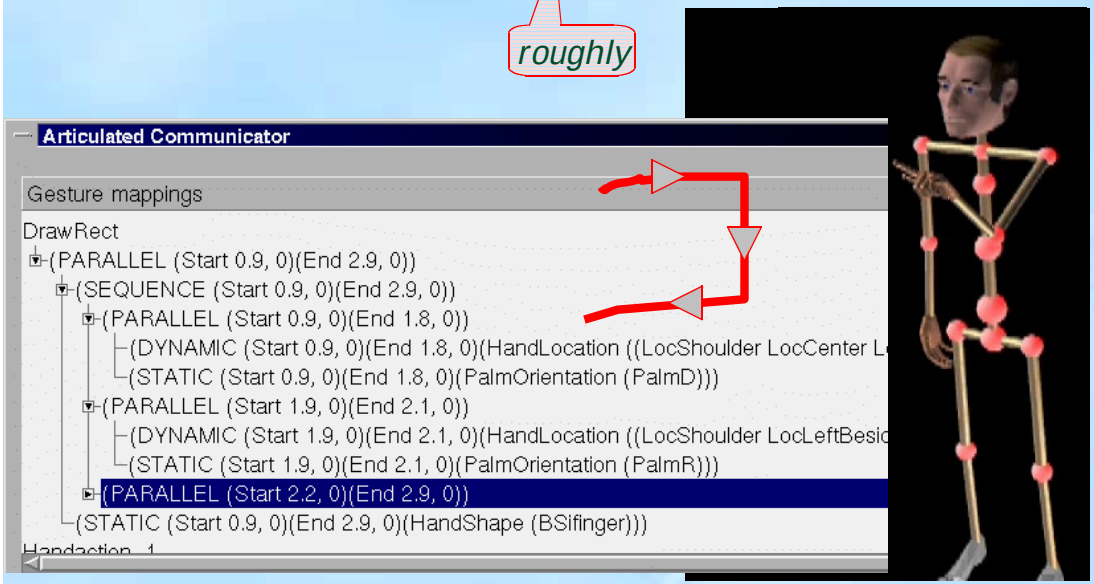
preparation (automatic) | stroke (specified) | retraction (automatic)

```
Pull-2
  (PARALLEL (Start 1.1)(End 1.41))
    (DYNAMIC (Start 1.1)(End 1.41)(HandLocation ((LocShoulder LocLeftBeside LocStretched)
                                                 (LocShoulder LocLeftBeside LocNear))) (Manner ((Peak 1.39))))
    (DYNAMIC (Start 1.1)(End 1.41)(HandShape ((BSflato)(BSfist)))(Manner ((Peak 1.39))))
    (STATIC (Start 1.1)(End 1.41)(PalmOrientation (PalmU)))
```

Universität Bielefeld

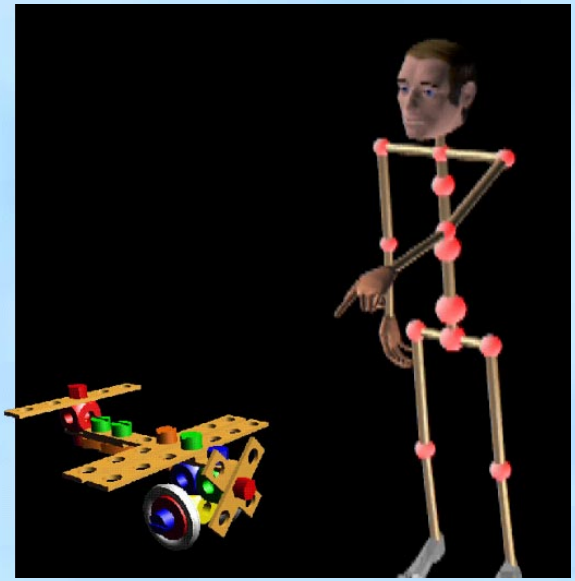
Outlining a rectangular shape

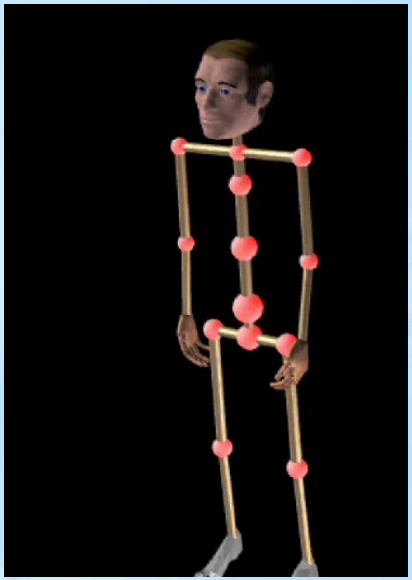
roughly



Universität Bielefeld

Pointing with and without beat



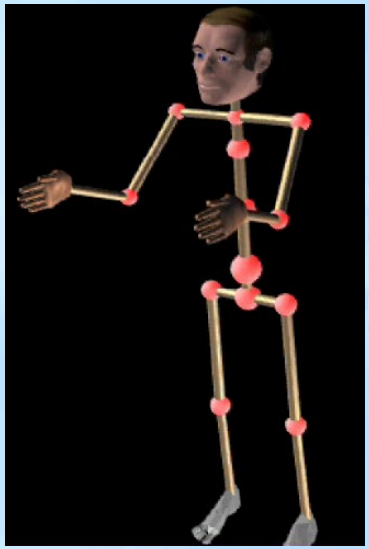


Universität Bielefeld

Iconic two-handed gesture

used to indicate the size or orientation of an object

e.g., referencing a single object in an assembly group

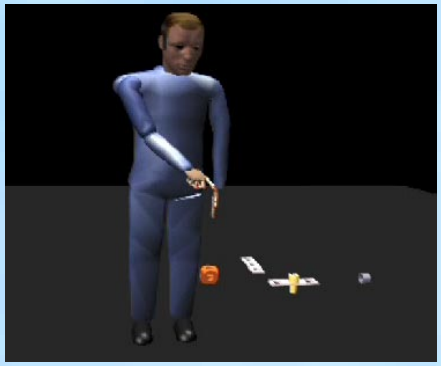


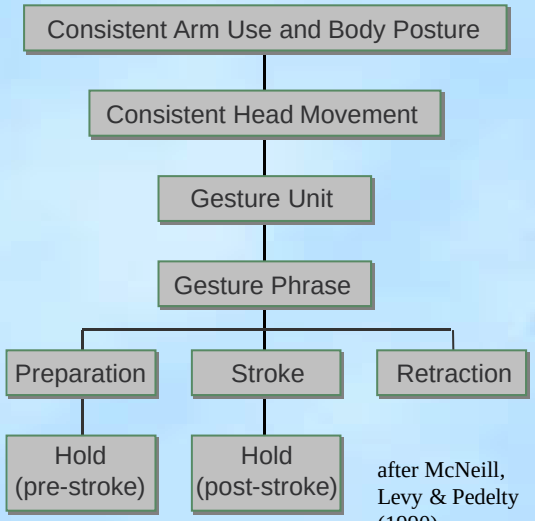
Universität Bielefeld

„MAX“ - bringing it all together

Multimodal Assembly EXpert

Work with Stefan Kopp, Bernhard Jung and master's students





after McNeill, Levy & Pedelty (1990)

Universität Bielefeld

Overview

- **Previous Work and Context**
- **Lifelike Gestures Synthesis**
- **Articulated Communicator**
- **Timing**
- **Outlook: Speech and Facial Gesture**

Universität Bielefeld

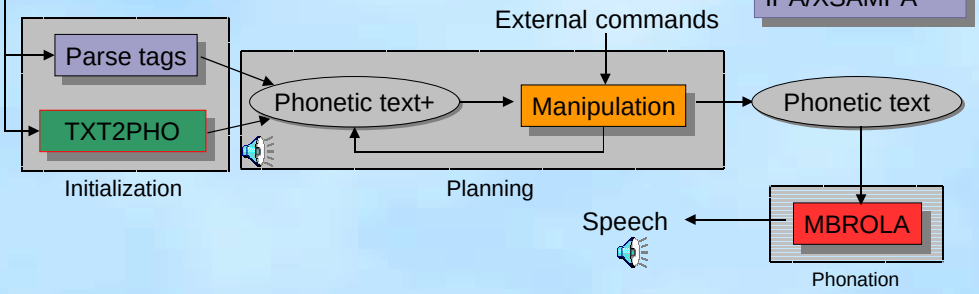
TTS for multimodality

- TXT2PHO (Uni Bonn) and MBROLA (Dutoit/Mons)
- SABLE tags for additional intonation commands

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

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

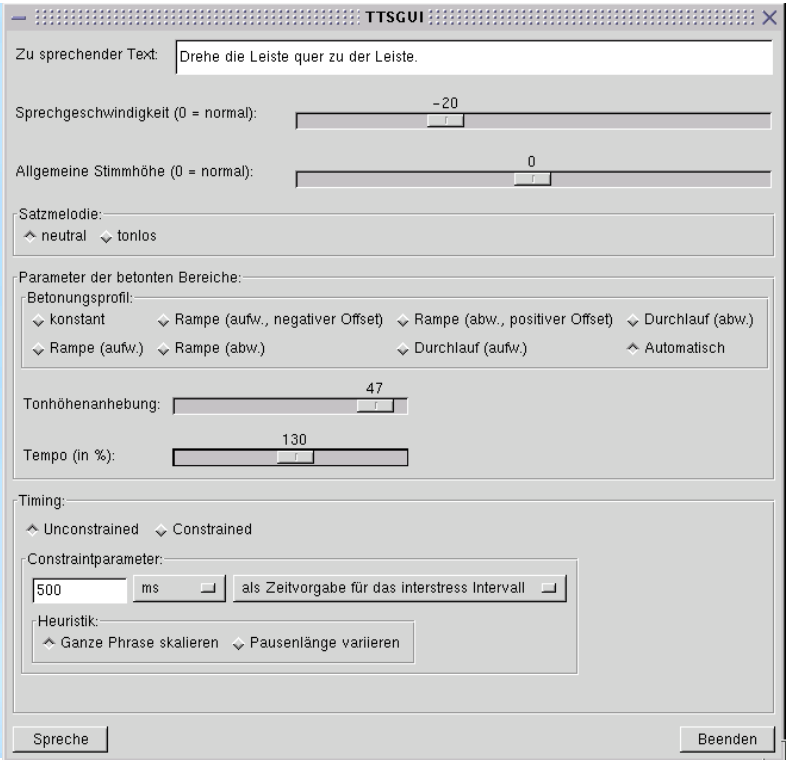
IPA/XSAMPA



Universität Bielefeld

Master's thesis
Dirk Stößel

In our text-to-speech system (TTS) a variety of prosodic functions can be controlled (pitch scaling, time scaling)



Universität Bielefeld

Conversational Agents: Aspects

Metaphor of face-to-face conversation in Interface design:

- mixed initiative dialogue
- nonverbal communication included
- (bodily) presence
- rules for transfer of control

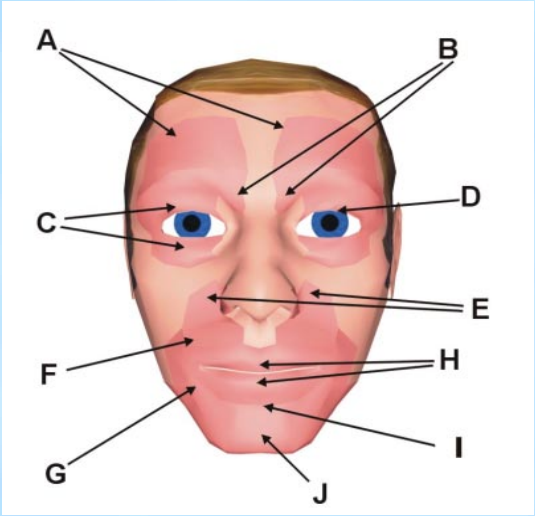
Some aspects of relevance:

- Personality
  - field of expertise
  - profile of interest
  - audiovisual appearance
- Performatives
- conversational functions
- Emotion

Much more challenging work to be done.
Have started to work on facial gesture....

Universität Bielefeld

Facial Animation



A:	Kopfhaubenmuskel - Stirnteil
B:	Augenbrauenrunzler
C:	Augenringmuskel
D:	Augenlid
E:	Kleiner Jochbeinmuskel, Nasenmuskel & Oberlippenheber
F:	Großer Jochbeinmuskel & Mundwinkelheber
G:	Mundwinkelherabzieher
H:	Ringmuskel des Mundes
I :	Unterlippenherabzieher
J:	Unterkiefer

Universität Bielefeld

Facial Gesture / Emotion ...




Will it be possible to model emotion and express it by way of facial gesture? For the six basic emotions: happiness and sadness, surprise, fear, disgust and anger there seem to be universal facial expressions. [after Ekman]

Universität Bielefeld

AI group Team

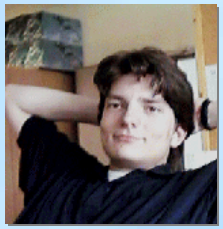

Bernhard Jung


Stefan Kopp

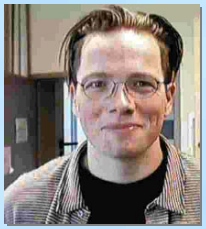

Peter Biermann

Technische Fakultät

Labor für Künstliche Intelligenz & Virtuelle Realität


Timo Sowa


Ian Voß


Marc Latoschik

Universität Bielefeld

The End

Universität Bielefeld

Lifelike Gesture Synthesis and Timing for Conversational Agents

Find Lab Showcase & papers...

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

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