Modeling gestures in the *MAX* system

Gesticons WS, Dec. 2003

Dr. Stefan Kopp

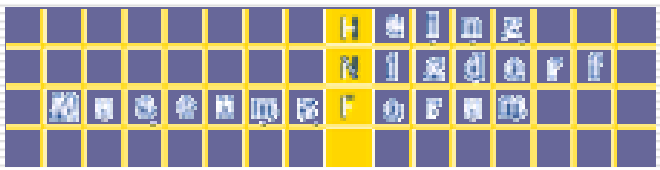
Artificial Intelligence Group
Faculty of Technology

University of Bielefeld

Max – employed in various contexts

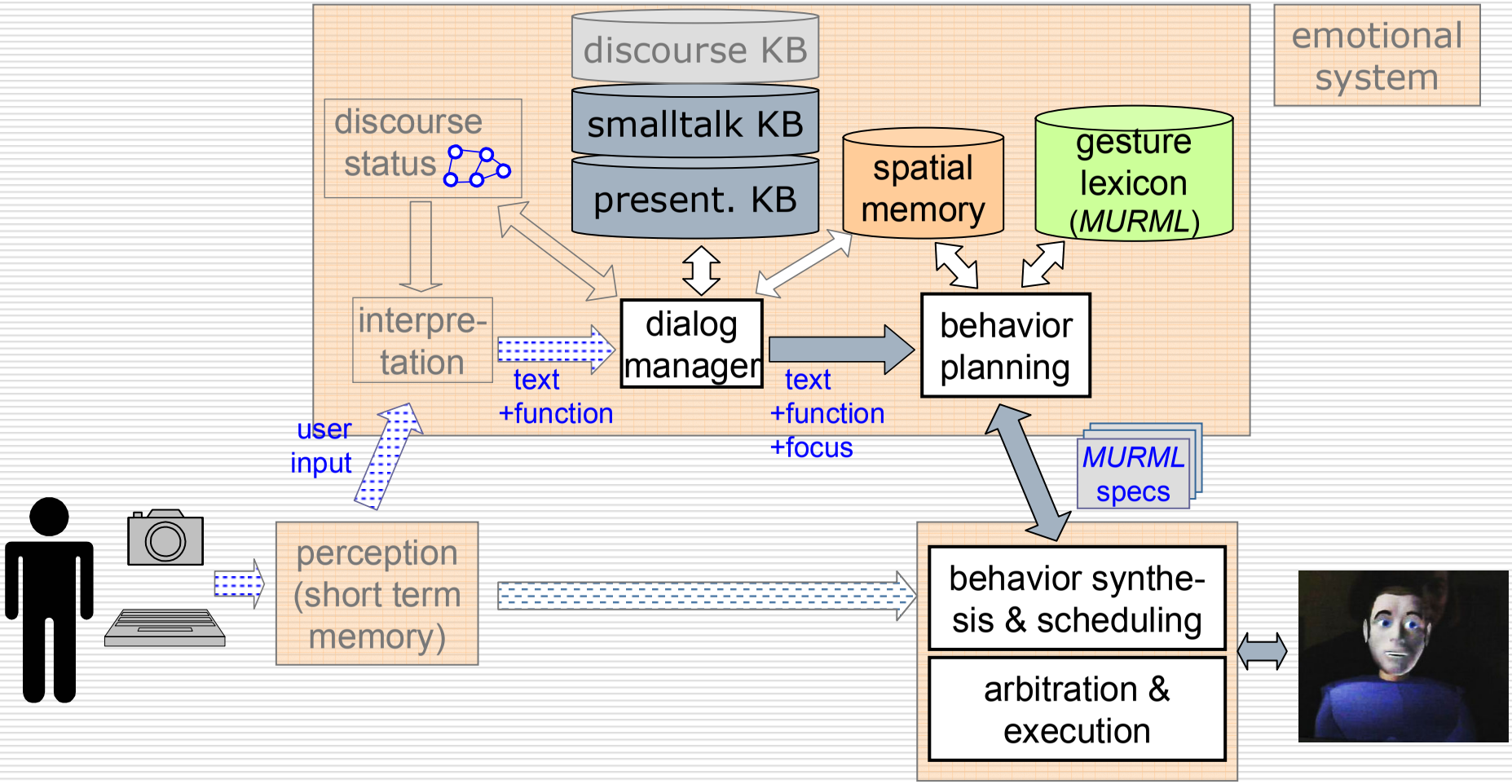
Multimodal Assembly eXpert

Presentation agent in the
Heinz Nixdorf MuseumsForum



from Jan.´04 part of the regular exhibition

Overall architecture



cf. *Rea* (Cassell et al.,'99), FXPAL (Churchill et al.,'99)

How gestures come into play…

Input to behavior planning:

```
<act function=„provide.content.self-reference“>
  But <focus> I am </focus> an artificial person.
</act>
```

- function: encodes pragmatic & semantic aspects of a conversational act and its focussed part (optional)
- focus: marks salient element, import to the speaker (rheme, constrast) → gesture placement hint
- extensible set of functions, hierarchically organized
  - *provide.content.admonition*
  - *provide.discourse.disagree*
  - *askFor.content.presentation*
  - *….*

Currently 218 functions

Lexicalization of gestures

```
<gesture id="raise-eyebrows" scope="face">
  <function name="signal-rejection"/>
  <function name="signal-sorry"/>
  <function name="signal-admonition"/>
  <constraints> .... </constraints>
</gesture>

<gesture id="head-shake" scope="head">
  <function name="signal-disagree"/>
  <constraints> .... </constraints>
```

each entry defines a behavior by bringing together a *form definition* with *communicative function(s)* it fulfills

function-form assignement is…

- ambiguous
- parameterised (*deictics* depend on spatial context)

gesture selection in two stages:

1. communicative function that supports the overall act's function
2. gesture that fulfills this function (→ gesture planner)

How to specify gesture form?

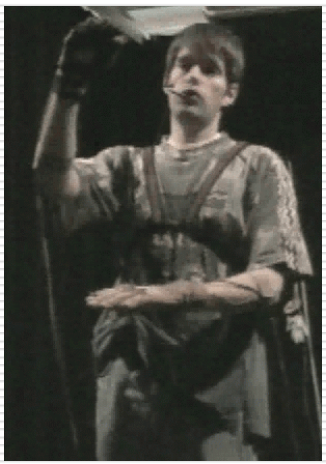
two basic approaches

1. feature-based representation
2. “control-based” representation

Feature-based represention of gesture form

specify the only *essential outer form features* (expressive phase)

- pros:
  - + directly describes overt gesture, easy to use
  - + tradition in sign language notation systems
  - + independent of body properties
  - + makes clear what of a gesture is mandatory and what is not
  - + facilitates gesture planning, especially for illustrative gestures



- cons:
  - - requires sophisticated animation planning & generation
  - - difficult to describe dynamic gestures, single-joint motions, self-touches

Feature-based representation in *MURML*

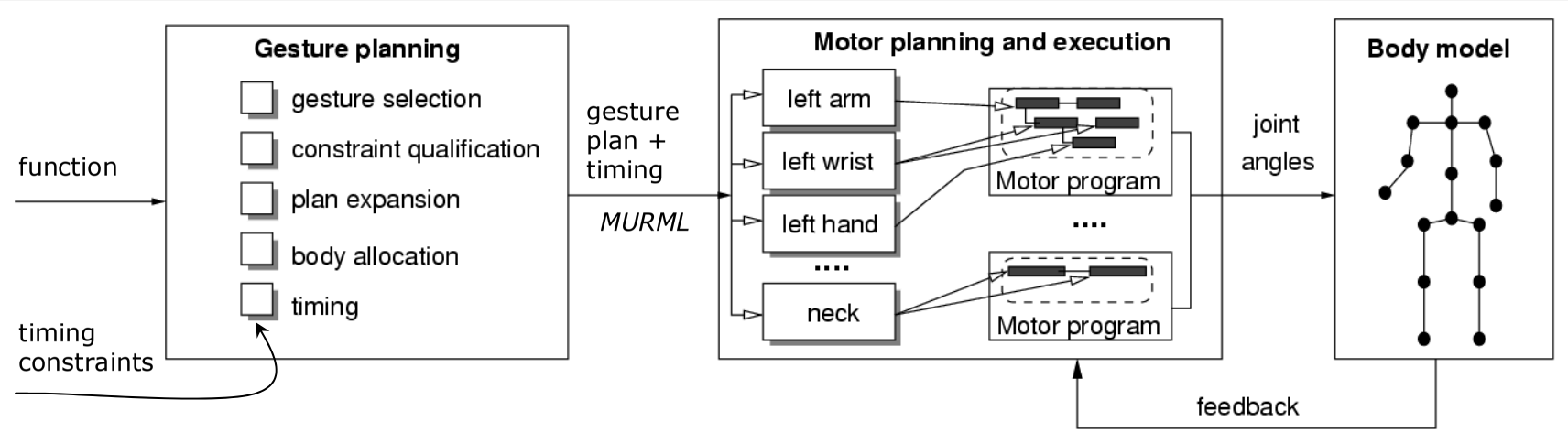
- *static* or *dynamic movement constraints*
- *hand shape*, *wrist location*, *extended finger orientation*, *palm orientation*
- numerical or symbolic values (*HamNoSys* repository)
- composition: *parallel*, *sequence*, *repetition*, *symmetry*

```
<constraints>
  <symmetrical dominant="right_arm" symmetry="SymMS">
    <parallel>
      <static slot="HandLocation" value="LocUpperChest
            LocCenterRight LocFar"/>
      <static slot="PalmOrientation" value="DirA"/>
      <static slot="ExtFingerOrientation" value="DirU"/>
      <static slot="HandShape" value="BSflat"/>
    </parallel>
  </symmetrical>
</constraints>
```

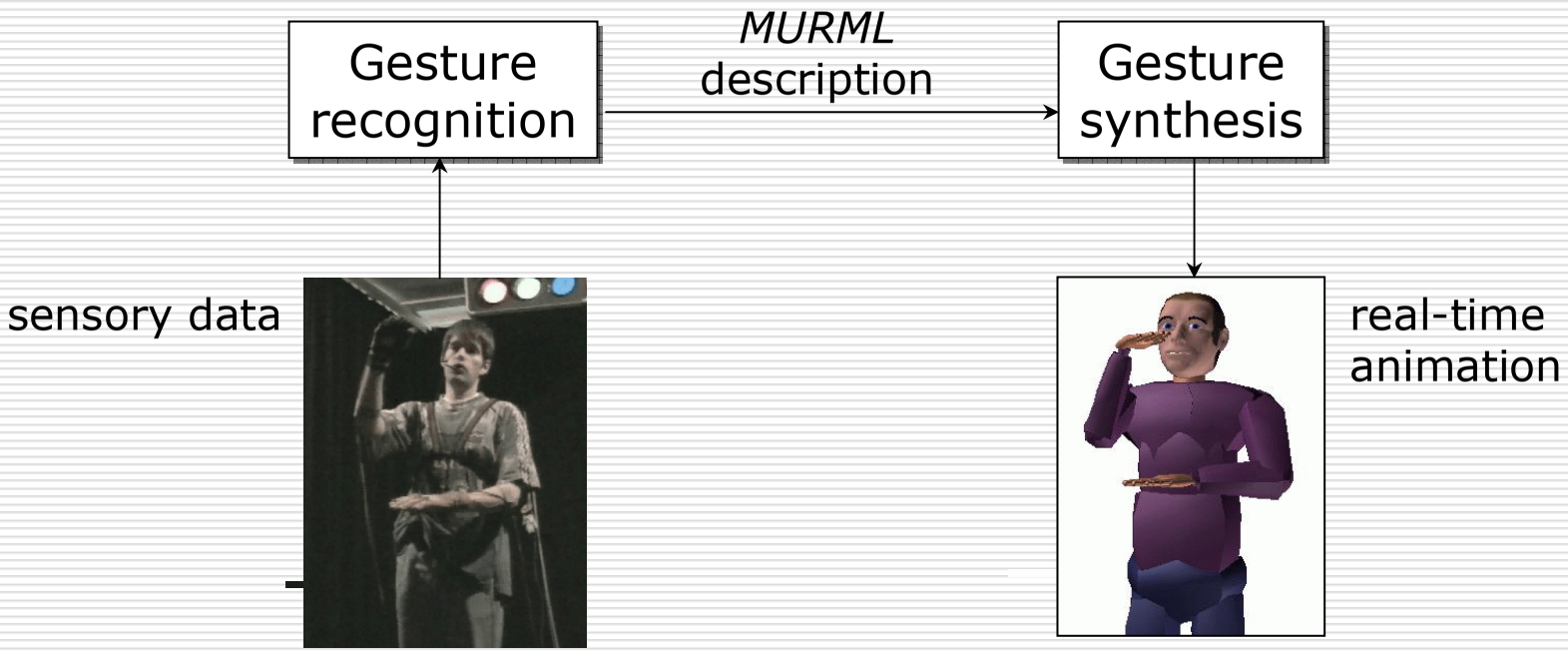
Max's gesture generation model

animations *created* on-the-fly from *MURML* definitions

- high-level gesture planning, low-level motor planning
- decomposition of motor control: concurrent local motor programs for combining motion generators, self-activation & mutual de-/activation
- embedded in incremental utterance production model



Application: gesture imitation in VR



„Control-based“ representation

describe motion in terms of *time-varying control parameters*

- usually *keyframes* in joint angles or postural features
- Pros:
  - + high motion quality achievable
  - + animating the character is simple and fast
- Cons:
  - - tedious definition of keyframe set, requires expert animator
  - - gesture hardly adaptable in a consistent way, additional info needed - at least about phase structure

MURML provides means of defining multi-phase *keyframe animations* for the skeleton and the face

Control-based representation in *MURML*

```
<constraints>
  <dynamic>
    <keyframing mode="spline" easescale="6.0" easeturningpoint="0.5">
      <phase>
        <frame ftime="0.2"><posture>
          (MUSCLE_BROW_LEFT 0.5) (MUSCLE_BROW_RIGHT 0.5)
          (MUSCLE_EYE_LEFT_ROUND_UP -0.5) (MUSCLE_EYE_RIGHT_ROUND_UP -0.5)
        </posture></frame>
        <frame ftime="0.8"><posture>
          (MUSCLE_BROW_LEFT 0.5) (MUSCLE_BROW_RIGHT 0.5)
          (MUSCLE_EYE_LEFT_ROUND_UP -0.5) (MUSCLE_EYE_RIGHT_ROUND_UP -0.5)
        </posture></frame>
        <frame ftime="1.0"><posture>
          (MUSCLE_BROW_LEFT 0) (MUSCLE_BROW_RIGHT 0)
          (MUSCLE_EYE_LEFT_ROUND_UP 0) (MUSCLE_EYE_RIGHT_ROUND_UP 0)
        </posture></frame>
      </phase>
    </keyframing>
  </dynamic>
</constraints>
```

We use both approaches: Currently, 33 gestures defined in a feature-based way and 21 (mostly symbolic and facial) gestures as keyframe animations

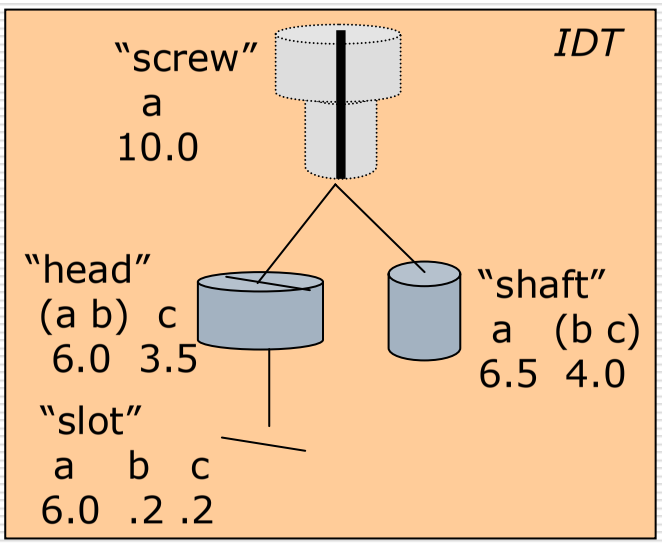
semantics – beyond functional modelling

"shape-related" iconic gestures

- *abstraction*
- *decomposition*
- *spatial coherence*

required:

- formalized notion of semantics (pictorial content)
- „sense-preserving" mapping to morphology (form)



Form-meaning mapping

associating *semantic features* (axes) with form features

#	Form feature	Axis type	Axis orientation	Axis degree
1	Linear move	1D	Movement direction	Mov. length
2	Circular move	2D	Movement plane	Diameter
3	Palms facing, flat	1D	Segment betw. palms	Palm distance
4	Palms facing, round	2D	Round plane	Palm distance
5	"Precision-grip"	1D	Aperture vector	Aperture
6	Flat hand	1D	Extended fing. orient.	(undefined)

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

skopp@techfak.uni-bielefeld.de