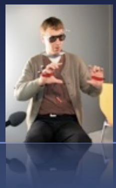
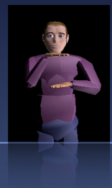


# GNetIc–Using Bayesian Decision Networks for Iconic Gesture Generation



*Kirsten Bergmann & Stefan Kopp*
*Sociable Agents Group, CITEC, Bielefeld University*

## Outline

1. The problem of generating speech-accompanying gestures
2. GNetIc gesture generation approach
   - Data corpus
   - Gesture formulation using decision networks
   - Structure learning
3. Evaluation results
4. Overall speech and gesture production architecture

## Human-like Expressiveness for Virtual Agents

Issues to decide on:
- Gesture (yes/no)
- Representation technique (placing, drawing etc.)
- Morphological features, e.g.
  - Handedness
  - Handshape
  - Orientation of palm and fingers
  - Movement trajectory

*„[...] why different gestures take the particular physical form they do is one of the most important yet largely unaddressed questions in gesture research“* (Bavelas et al. 2008, p. 499)

## Multiple Influences on a Gesture‘s Form

Discourse Context (Bergmann & Kopp, 2009)
- Communicative goal
- Information structure

Previous Gesture Features
- Catchments (McNeill, 2005)

Interlocutor‘s Communicative Behavior
- Gestural mimicry (Kimbara, 2006)

Linguistic Context
- Linguistic information packaging (Kita & Özyürek, 2003)

## Experimental Setting

- Direction giving and sight descriptions for a VR town
- Simplified objects to allow determination and control of message content (shape features, level of detail)

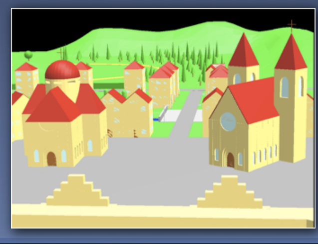
- Audio- and video recordings (3 camera views)
- Motion tracking of hand and head movements
- 25 dyads
  - ~350 min
  - 39.435 words
  - 4.961 iconic/deictic gestures

## SaGA corpus

### 1. Gesture annotation

- Gesture segmentation
- Gesture classification
- Gesture morphology

### 2. Referent features

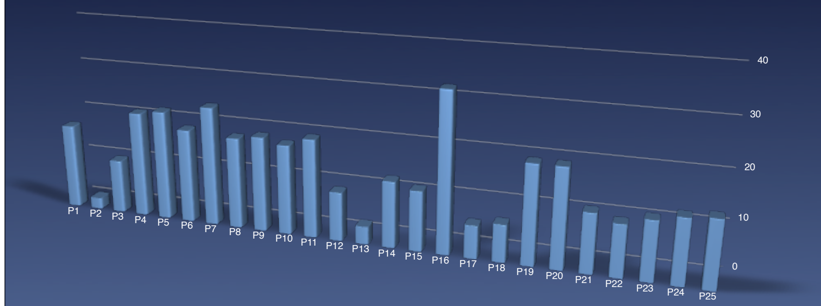


### 3. Gesture context

- Speech
  - Words
  - POS
  - Syntactic constructions
- Discourse Context
  - Information Structure
  - Communicative Goals
  - Dialogue Acts

## Idiosyncrasy in Gesture Use

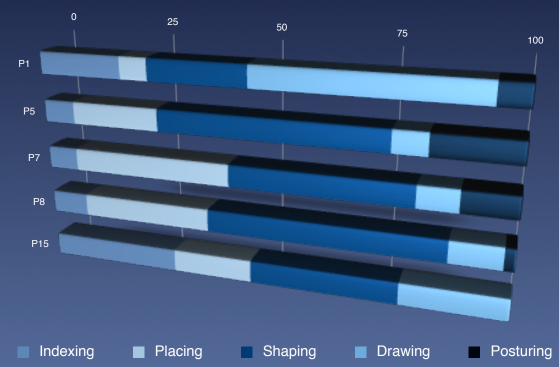
# Gestures per minute

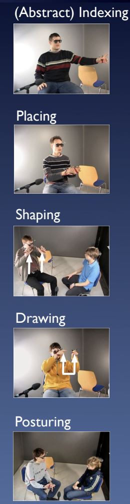



n = 25, M = 15.68, min = 2.34, max = 32.83, SD = 7.08

## Idiosyncrasy in Gesture Use

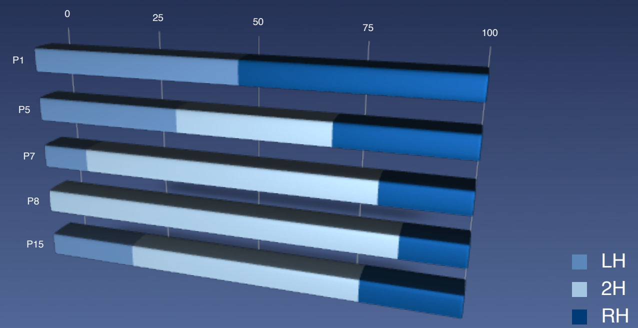
Representation Techniques



# Idiosyncrasy in Gesture Use

Handedness in placing gestures




# Results: Systematicity vs. Idiosyncrasy

Individual patterns

- Number of gestures
- Use of representation techniques
- Morphological gesture features

Bergmann & Kopp (2009), Kopp et al. (2007)

Inter-subjective patterns

- Correlation of representation technique and referent features
- Correlation of communicative intention and representation technique
- Correlations between visuo-spatial referent features and morphological gesture features

How to account for both in a computational model?

# Gesture Generation Approach

Gesture Net for Iconic Gestures (GNetIc)

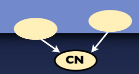
Bayesian Decision Networks (Howard & Matheson, 2005)

- Representation of sequential decision problems
- Combination of probabilistic and rule-based decision-making
- Decisions represented as
  - Chance nodes
    for variables found to be highly idiosyncratic
  - Decision nodes
    for variables with intersubjective correlations

*HUGIN toolkit (Madsen et al., 2005)*

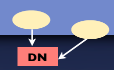
# Bayesian Decision Networks

**Chance Nodes**




- Learned from corpus data
  - For each individual speaker
  - For the whole corpus
- Structure learning
  - Score-based Learning
  - Constraint-based Learning
- Parameter estimation
  - Estimation of the distribution parameters
  - EM parameter learning (Lauritzen, 1995)

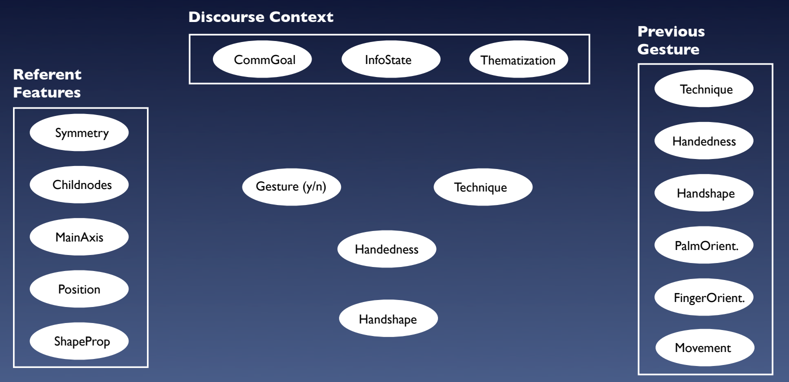
**Decision Nodes**




- Definition of if-then rules for
  - Palm and finger orientation
  - Movement features
- Accounting for inter-subjective correlations
- Chance nodes and referent features as predecessors
- 50-100 rules per node, e.g.:

```
if (and (Gesture="true", Technique="drawing",
Handedness="rh", Handshape="ASL-G",
ShapeProp="round2d"), "MR>MD>ML>MU").
```

# Building Networks: Structure Learning




# Structure Learning

## Score-based

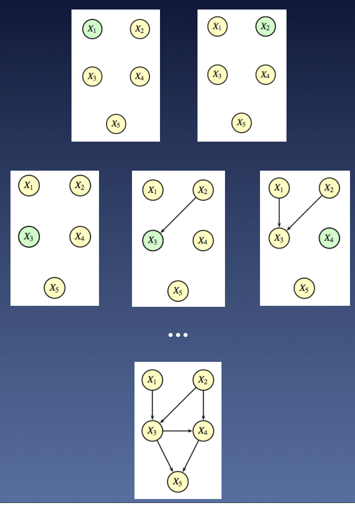
- Definition of a global measure (score) to evaluate a BN as a function of data
  - Bayesian score (Cooper & Herskovitz, 1992)
  - Bayesian Information Criterion (BIC; Schwarz, 1978)
- Searching the space of possible BNs to find best scored BN
  - Local search algorithms
  - Global search algorithms

## Constraint-based

- Conditional independence tests on database
- Building a directed acyclic graph in agreement with statistical test results

# Score-based Structure Learning

- **K2 algorithm** (Cooper & Herskovitz, 1992)
  - Greedy search for sequential decision problems
  - Initially each node has no parents
  - Algorithm incrementally adds that parent whose addition increases the resulting score most
  - Stops when addition of no single parent can increase the score

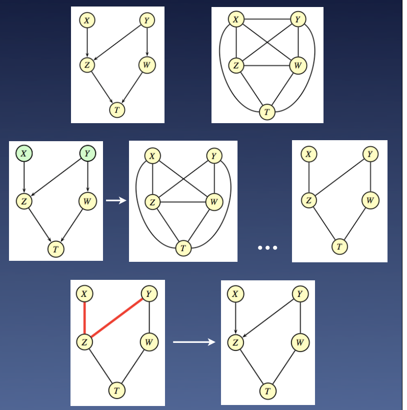


# Score-based Structure Learning

- **Metropolis-Hastings** (Madigan & York, 1995)
  - Global search method
  - Markov chain Monte Carlo algorithm (MCMC)
  - Construction of a Markov chain whose state space is a set of DAGs
  - Sampling from this chain for „long enough“ (burn-in time) to ensure it has converged to its stationary distribution
  - Not deterministic

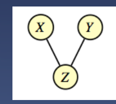
# Constraint-based Structure Learning

- **PC algorithm**
  (Spirtes & Glymour, 1991)
  - Starts from a complete, undirected graph
  - Deletes edges based on conditional independence decisions
  - Statistical tests for all pairs of variables (→ skeleton graph)
  - Directions are enforced (collider identification) ensuring that no direct cycles occur
  - Remaining undirected links are directed randomly (definition of constraints for counter-intuitive links)

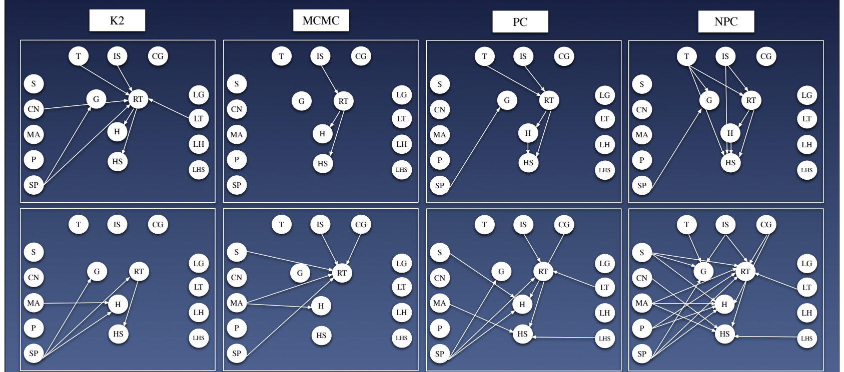


# Constraint-based Structure Learning

- **NPC algorithm**
  (Steck & Tresp, 1999)
  - Extension of PC algorithm
  - Especially for small data sets
  - Necessary Path Condition:
    - If X and Y are independent conditional on a set S, there must exist paths between X and every Z in S, and between Y and every Z in S. Otherwise the inclusion of Z in S is unexplained.
    - A number of links are required to be present in the graph

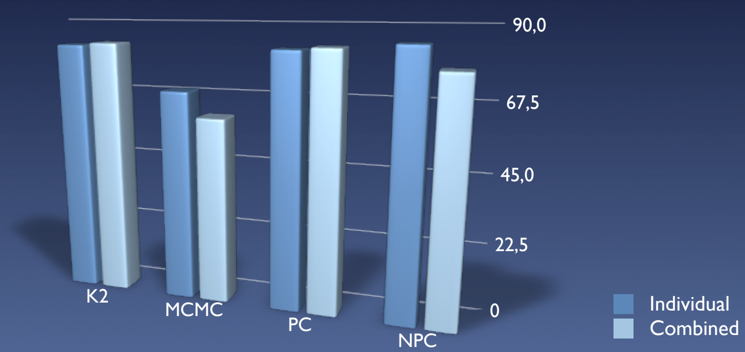



# Structure Learning: Results

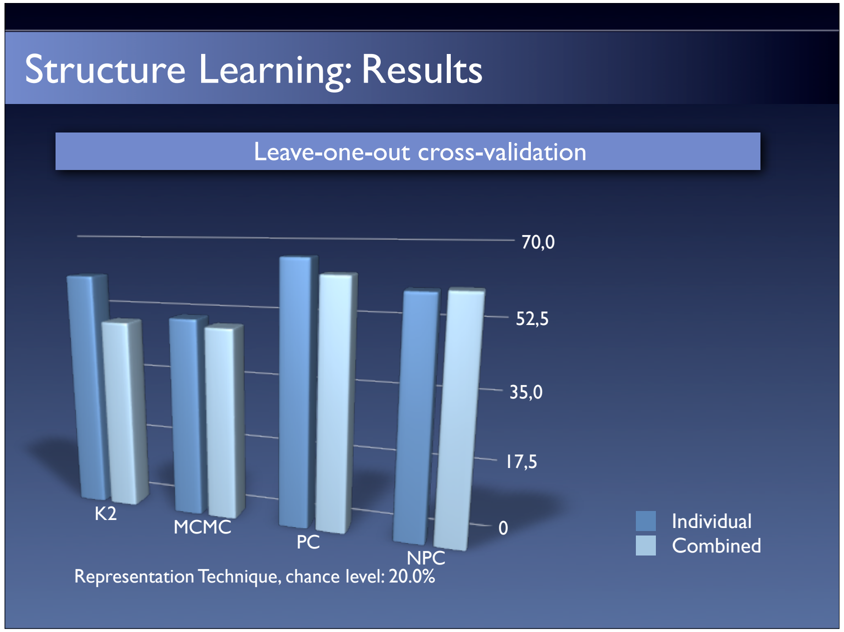


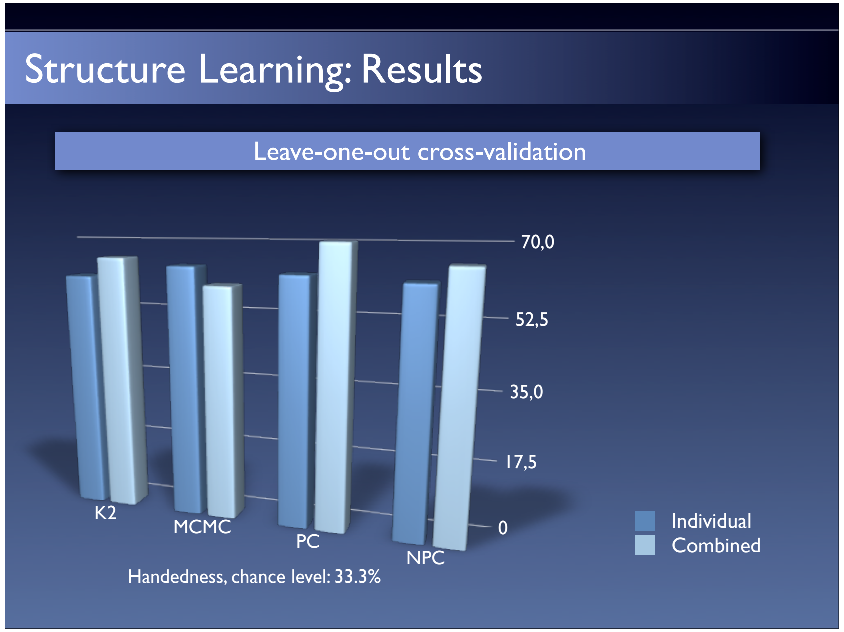

# Structure Learning: Results

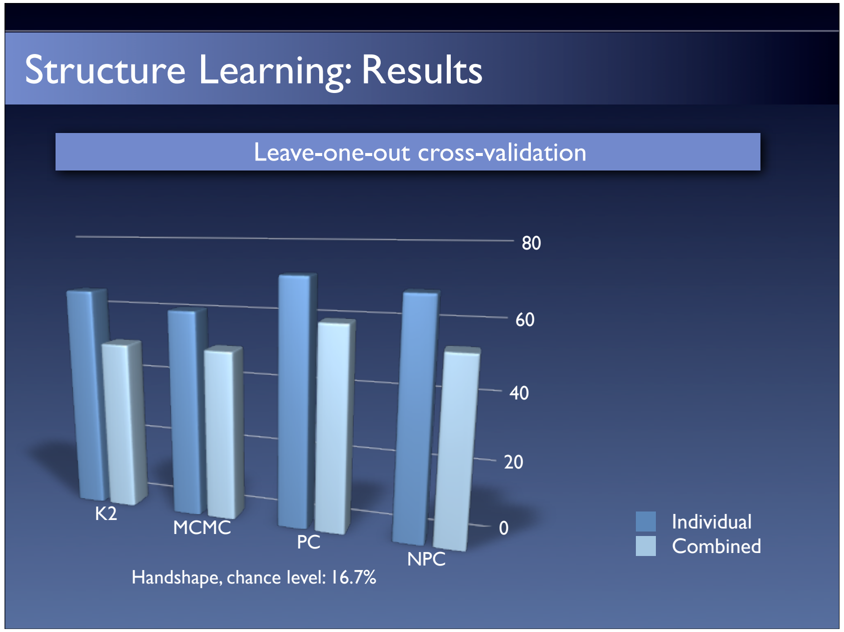
Leave-one-out cross-validation

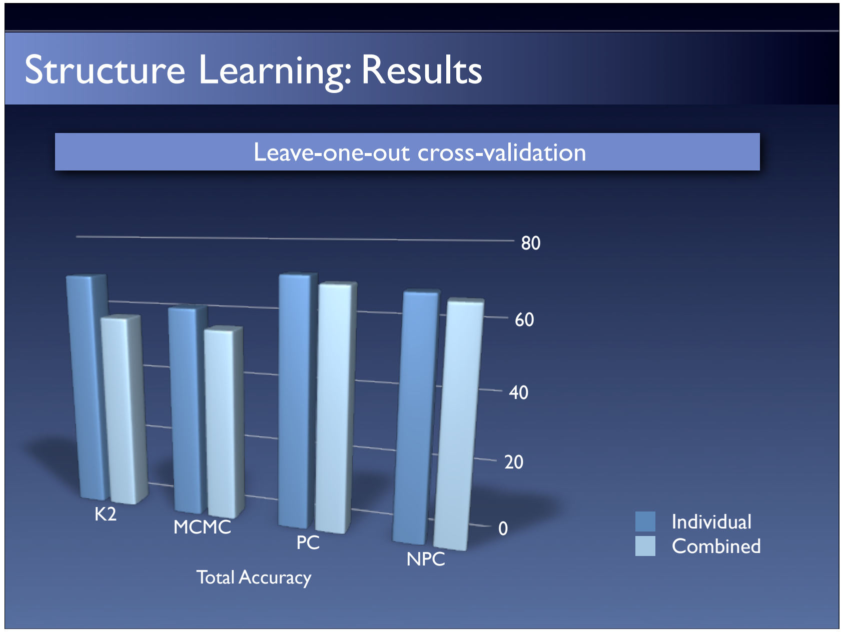



Gesture (yes/no), chance level: 50%


Structure Learning: Results
Leave-one-out cross-validation
70,0
52,5
35,0
17,5
0
K2
MCMC
PC
NPC
Individual
Combined
Representation Technique, chance level: 20.0%

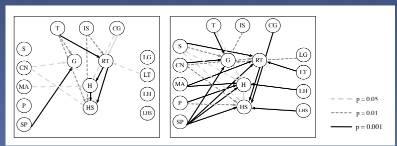

Structure Learning: Results
Leave-one-out cross-validation
70,0
52,5
35,0
17,5
0
K2
MCMC
PC
NPC
Individual
Combined
Handedness, chance level: 33.3%


Structure Learning: Results
Leave-one-out cross-validation
80
60
40
20
0
K2
MCMC
PC
NPC
Individual
Combined
Handshape, chance level: 16.7%

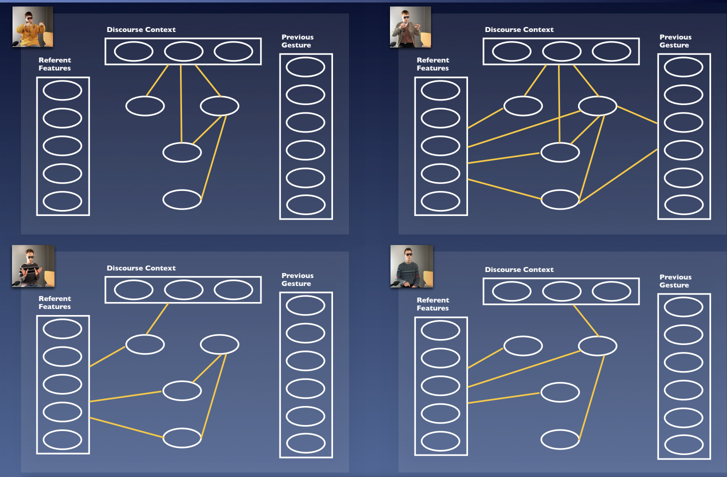

Structure Learning: Results
Leave-one-out cross-validation
80
60
40
20
0
K2
MCMC
PC
NPC
Individual
Combined
Total Accuracy

## Summary of Results

- Accuracy values:
  - clearly above chance level for all algorithms
  - constraint-based algorithms at least as good as score-based algorithms
  - individual BNs mostly equal/better than combined BNs
- Network structures differ (inconsistent edges)
  - Score-based algorithms
    - consider a *global* measure for entire network
  - Constraint-based algorithms
    - do not consider structure of the network as a whole (*local*)
    - Variation of significance level allows to consider link strength

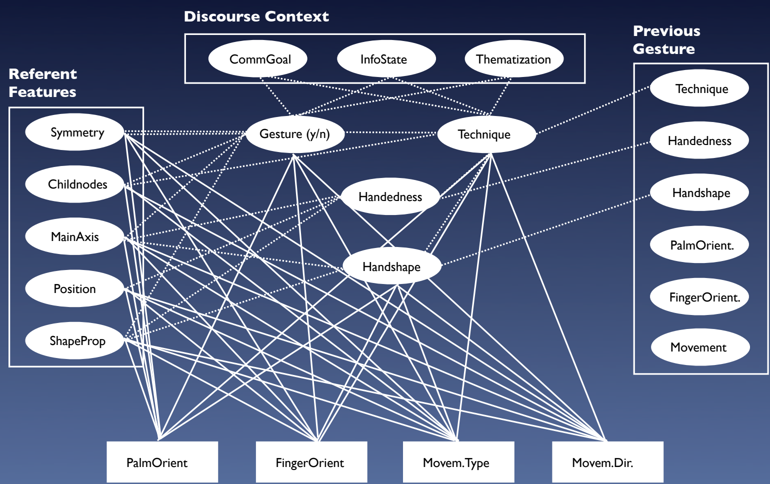



## Individual Speaker Networks



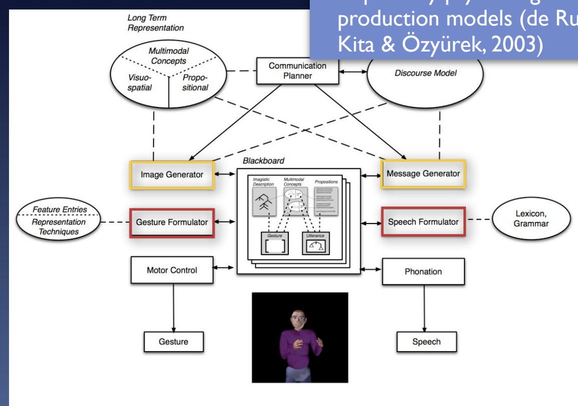

Idiosyncrasy in the production process of iconic gestures

## Building Networks: structure Learning




## Overall computational model

Inspired by psycholinguistic production models (de Ruiter, 2000; Kita & Özyürek, 2003)

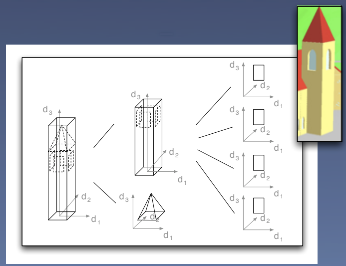



Multimodal content planning and formulation of speech and gesture run in parallel and interact on multimodal working memory

# Content Representation

## Computational Imagery

- Object shape represented as **Imagistic Description Trees** (Sowa & Wachsmuth, 2005)
  - Hierarchical structure
  - Extents in different dimensions
  - Possible underspecification



## Linguistic spatial representation

- Knowledge drawn upon by speech formulation
- Logical formulae based on a formal ontology of domain knowledge
- Partly reflects imagistic knowledge

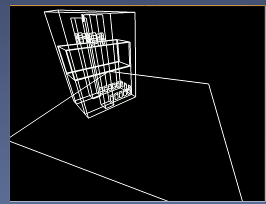
```
shared (entity (church-5, single)).
shared (instance_of (church-5, church)).
private (entity (window-3, single)).
private (instance_of (window-3, window)).
private (property (window-3, round)).
private (relpos (window-3, middle)).
private (entity (churchtower-1, single)).
private (instance_of (churchtower-1, tower)).
private (relpos (churchtower-1, left)).
private (part_of (church-5, churchtower-1)).
private (entity (churchtower-2, single)).
private (instance_of (churchtower-2, tower)).
private (relpos (churchtower-2, right)).
private (part_of (church-5, churchtower-2)).
...
```

# Generation Example

Communicative goal
DescrConstruct (church-5, churchtower1, churchtower-2)

## Image Generator

- Activates relevant imagistic descriptions
- Adopts spatial perspective



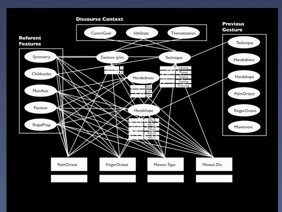
## Message Generator

- Collects propositions from long-term memory

# Generation Example

## Gesture Formulator

- Bayesian Decision Network to fill gesture feature matrix

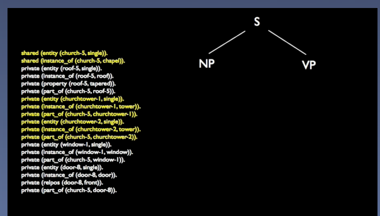



Gesture matrix

LOC: Periphery Left
TRAJ: Linear
MOVEMENT DIR: Down
HANDSHAPE: ASL-bent-5
PALM DIR: Down

## Speech Formulator

- Microplanning with SPUD (Stone et al., 2003)
  - Grammar-based microplanner using a Lexicalized Tree Adjoining Grammar (LTAG)




# Surface Realization

Gesture matrix

LOC: Periphery Left
TRAJ: Linear
MOVEMENT DIR: Down
HANDSHAPE: ASL-bent-5
PALM DIR: Down

- MAX/ACE (Kopp & Wachsmuth, 2004)
  - On-the-fly speech synthesis and movement planning
  - Scheduling and co-articulation of speech and gestures, incremental chunks (intonation phrase + gesture phrase)

```
<constraints>
  <parallel>
    <static slot="HandShape" value="BSifinger"/>
    <repeat_alt times="3">
      <dynamic slot="HandLocation">
       <dynamicElement type="curve">
         <value type="start"  name="LocShlder LocCentLeft LocFar"/>
         <value type="end"    name="LocShlder LocCentLeft LocNorm"/>
         <value type="normal" name="DirU"/>
         <value type="shape"  name="LeftC"/>
         <value type="extension" name="0.6"/>
       </dynamicElement>
      </dynamic>
    </repeat_alt>
  </parallel>
</constraints>
```

# Modeling Results