Honda Research Institute EU

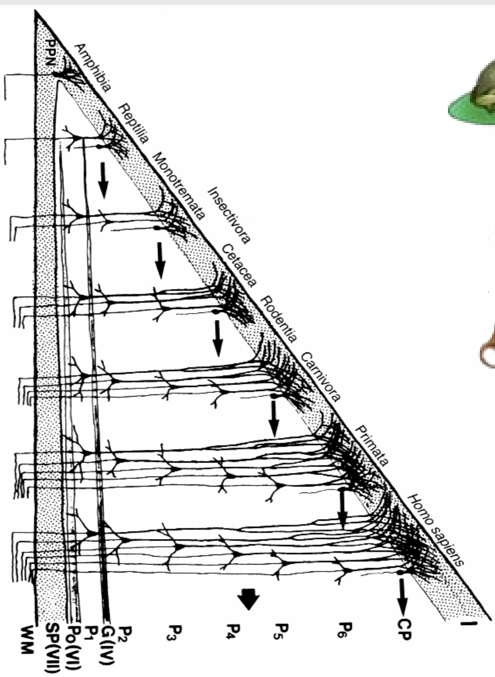
The Systems View on Interactive Online Multimodal Association for Internal Concept Building

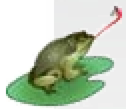
C.Goerick

innovation through science

Biologically Inspired Incremental Design

The system that bootstraps the development is so complex, that it demands itself incremental building.





Temporary reflex inhibition

Conditional and Instrumental learning

Control for tool and symbol usage

Behavioural Robotics
Reactive: sense-act

missing incremental steps ?

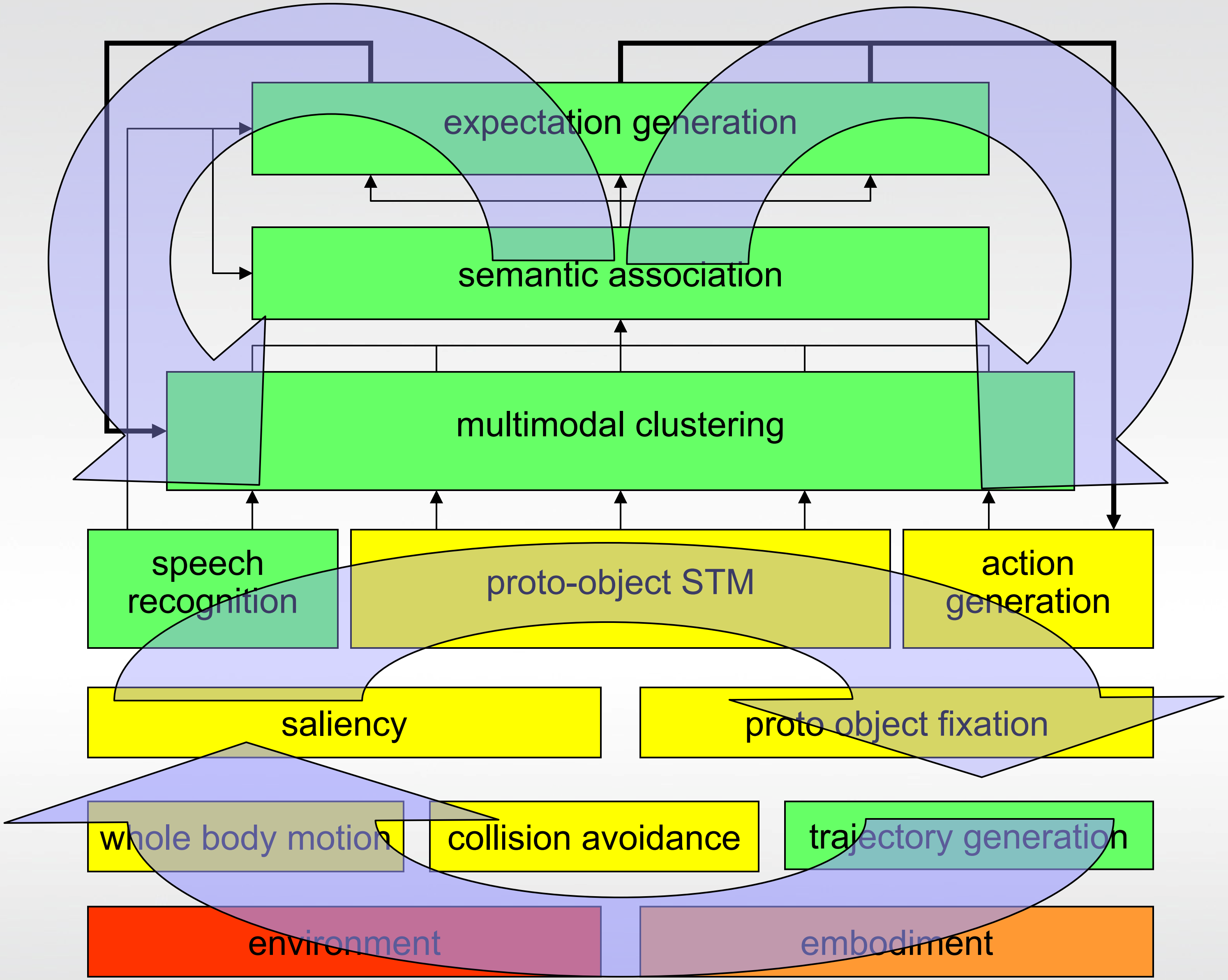
Biology: **Multiple control layers with different degree of prediction and optimization**

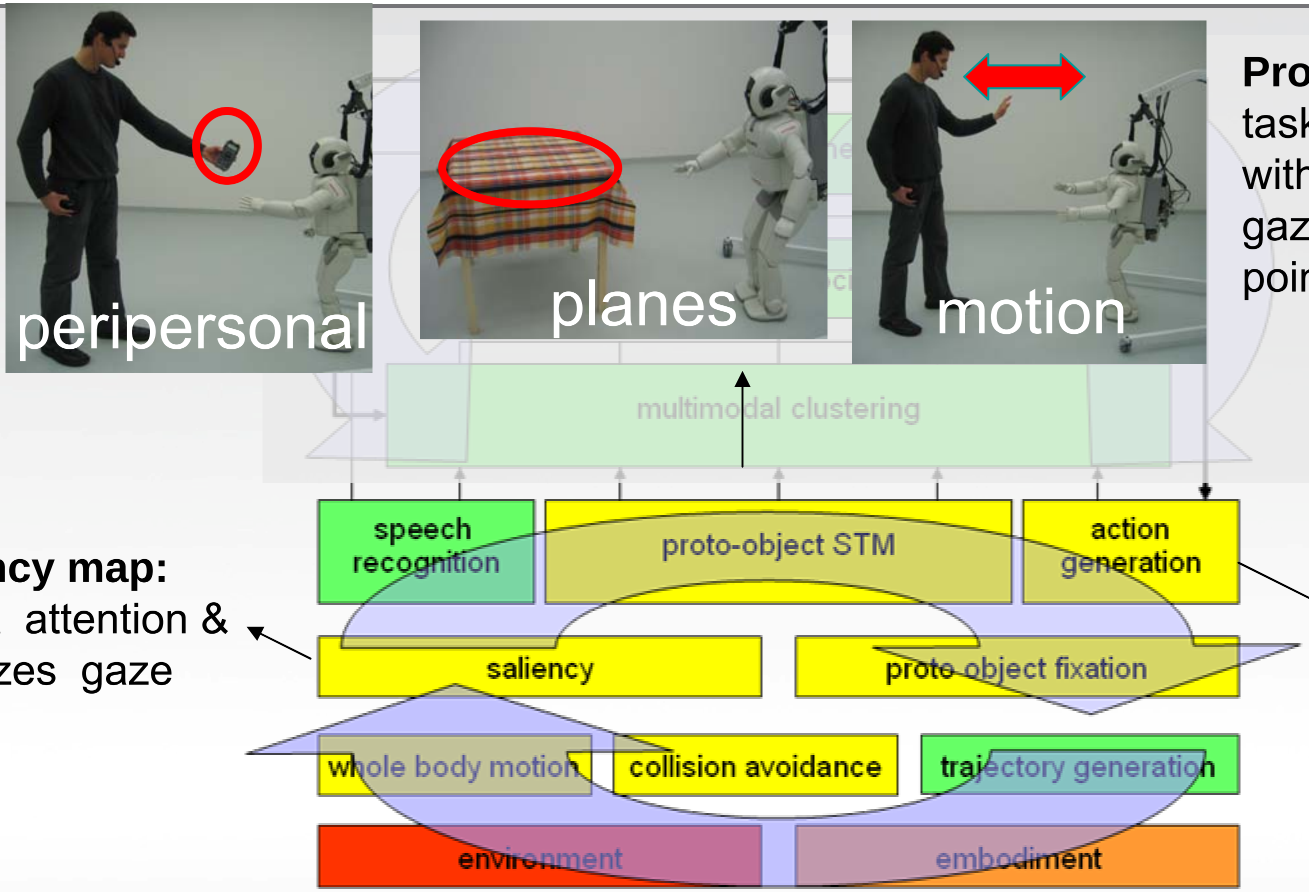
missing incremental steps ?

Classical AI
Plan: sense-plan-act

Subgoal:

internal concept building, interactive acquisition of multi-modal associations as grounding of concepts and symbols


expectation generation
semantic association
multimodal clustering
speech recognition
proto-object STM
action generation
saliency
proto object fixation
whole body motion
collision avoidance
trajectory generation
environment
embodiment



Proto-Object: bootstraps task-unspecific interaction with the environment: gazing, approaching, pointing

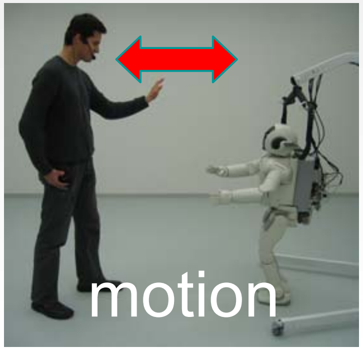
Saliency map: attract attention & stabilizes gaze

Arbiter: competitive selection of (paralel) behaviours

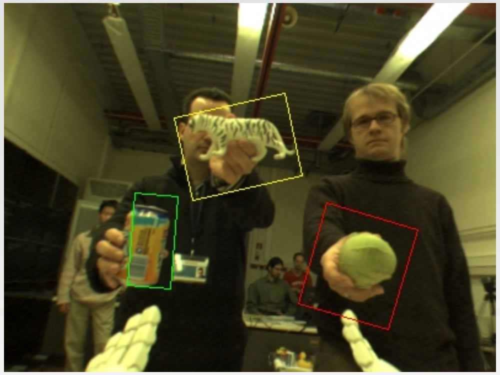
- Reactive Part
  - keeps system in favorable interaction, can be modulated i.e. cognitively overwritten
  - Solves several invariance problems
  - Provides continuously homogeneously represented percepts and propriocepts of the environment and the own body
  - Works without planning or any cognitive commands from higher levels

Representations in the reactive part

- Saliency maps
  - Activation maps in head pan-tilt coordinates
  - Based on visual contrast or strong auditory signals
  - Useful for first interactions with a system
- Proto-objects
  - Currently visually defined
  - Pre-identity references to parts of the outside world
  - Task unspecific
  - 3D description
  - Allows for first rough interaction based on segregated entities
  - Basis for possible task driven refinement
  - Cues:
    - Objects in peripersonal space around the robot
    - Homogeneously moving areas in the visual field
    - Planar supporting surfaces
- Auditory streams: MFCC or RASTA & HIST features
- Activations of reactive behaviors
  - Activity status of possibly parallel executed reactive behaviors

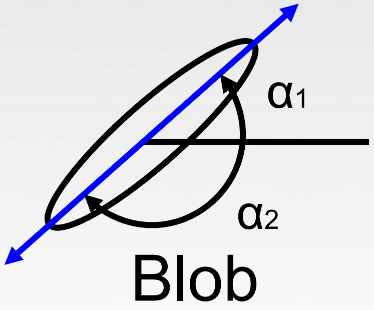


More on Proto Objects

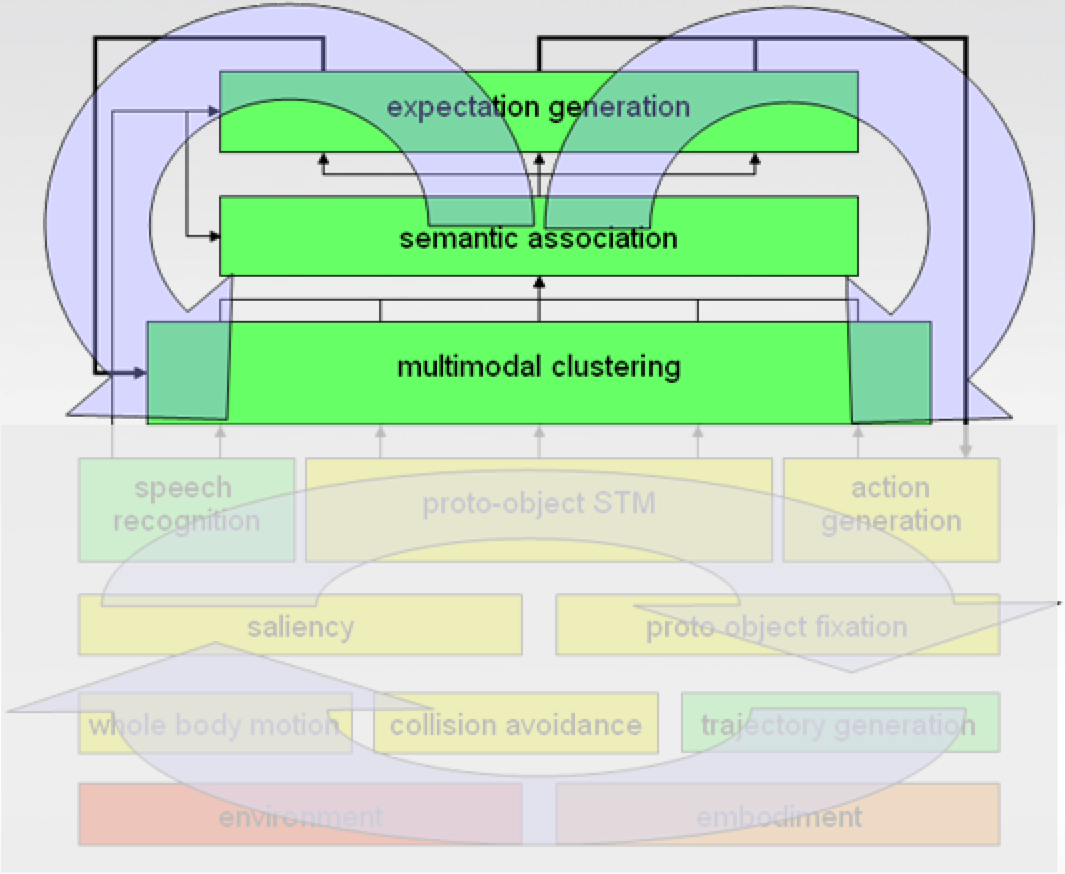


- Provide enough information to help following behavior:
  - position, orientation
  - features (size, strength, color, pitch, ...)
- No information is needed for:
  - recognition
  - model dependent features
- Proto objects are temporarily stabilized before further use in form of:
  - tracking
  - noise reduction
  - Memorization
- All further properties are derived from the Proto-Object representations

- Status, timestamp, ID, type.
- Optional position, orientations, and size in one or more coordinate systems.
- Orientations:

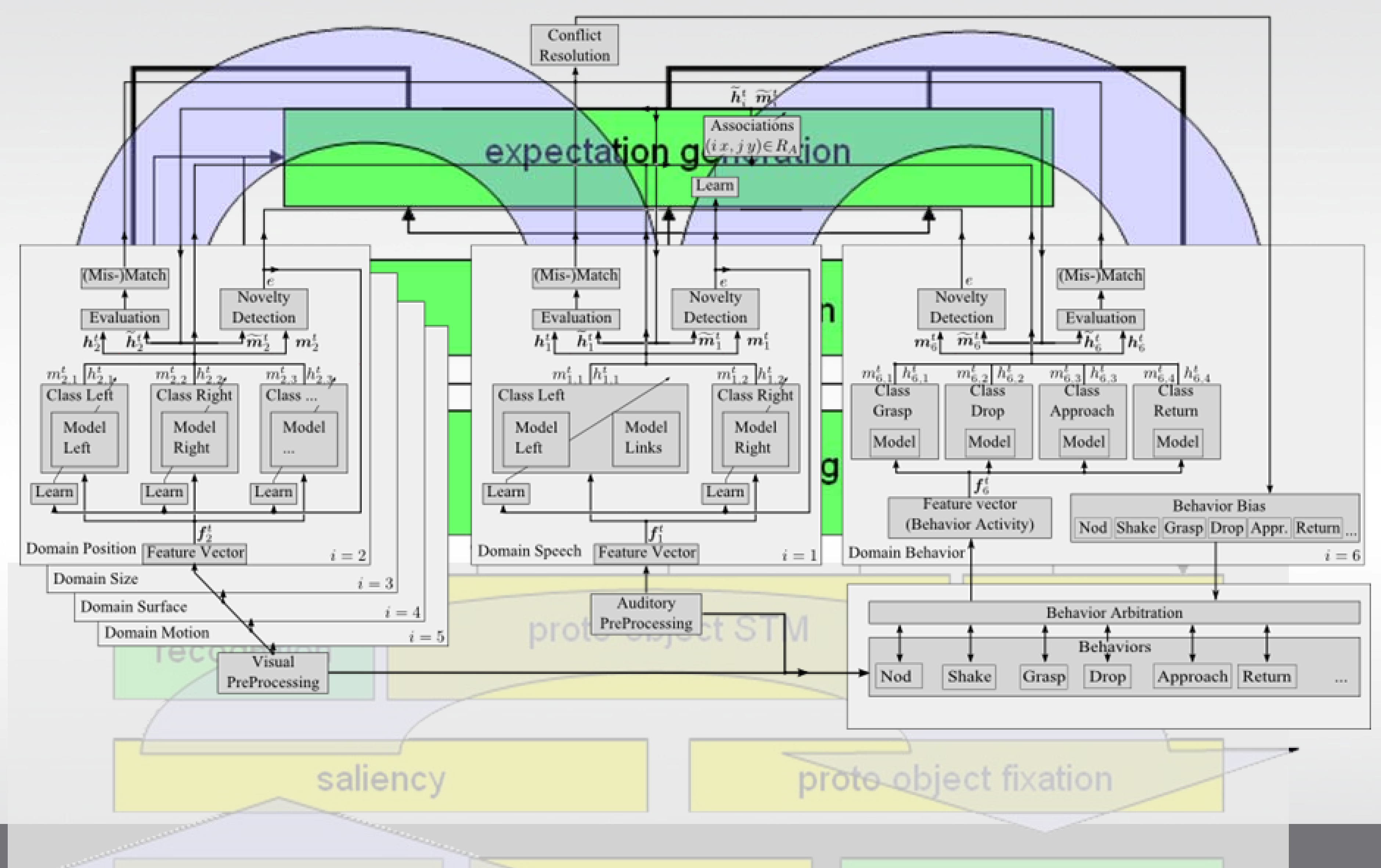


- STM maps ID of incoming PO to an existing or new PO. Orientations are disambiguated. Output:
  - Direct matches
  - Stabilized / filtered with prediction.

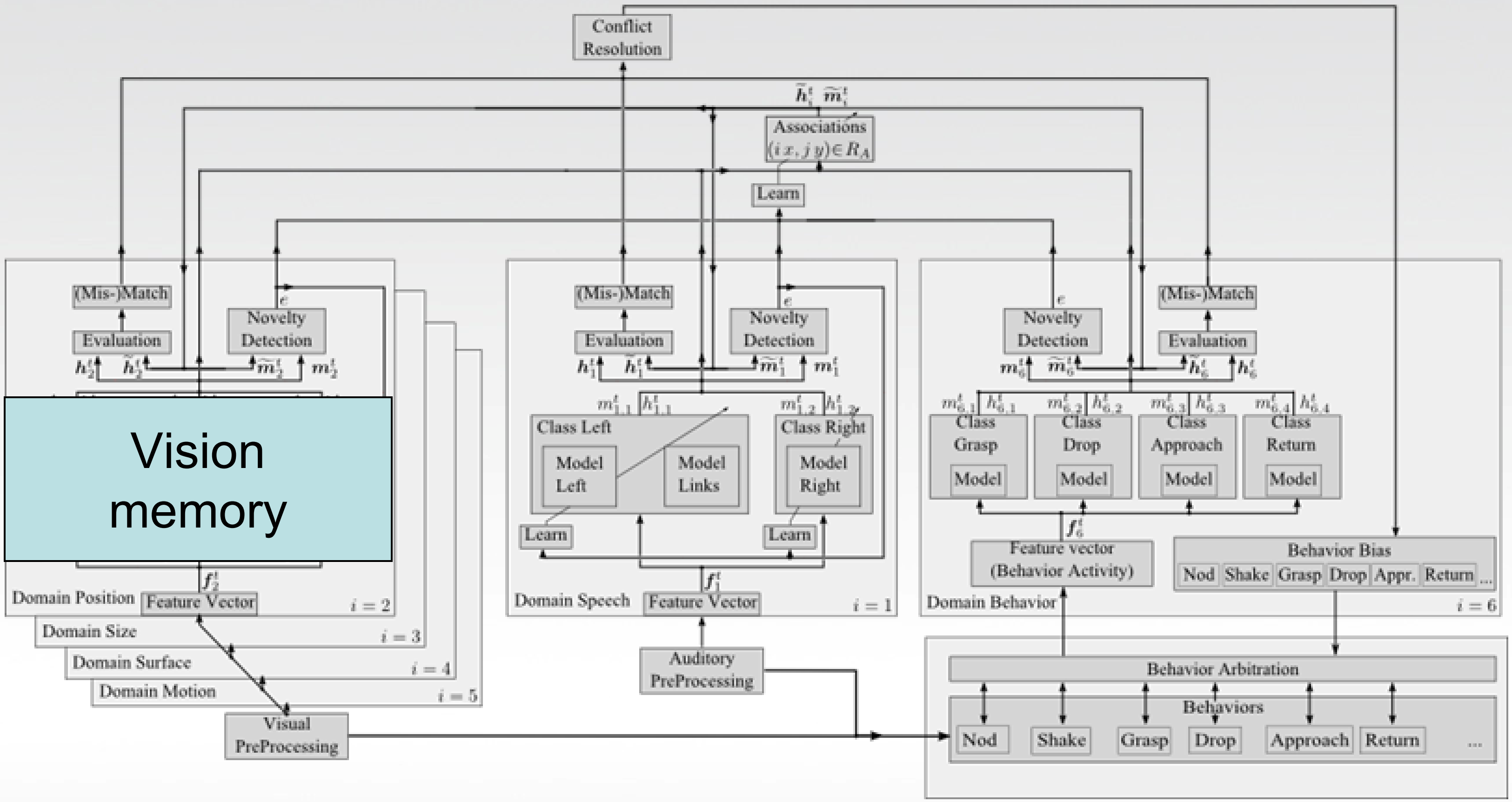


- Abtraction layer
  - Learns classes of the extero- and propriocepts
  - Learns associations between the classes
  - Organizes expectation driven behavior

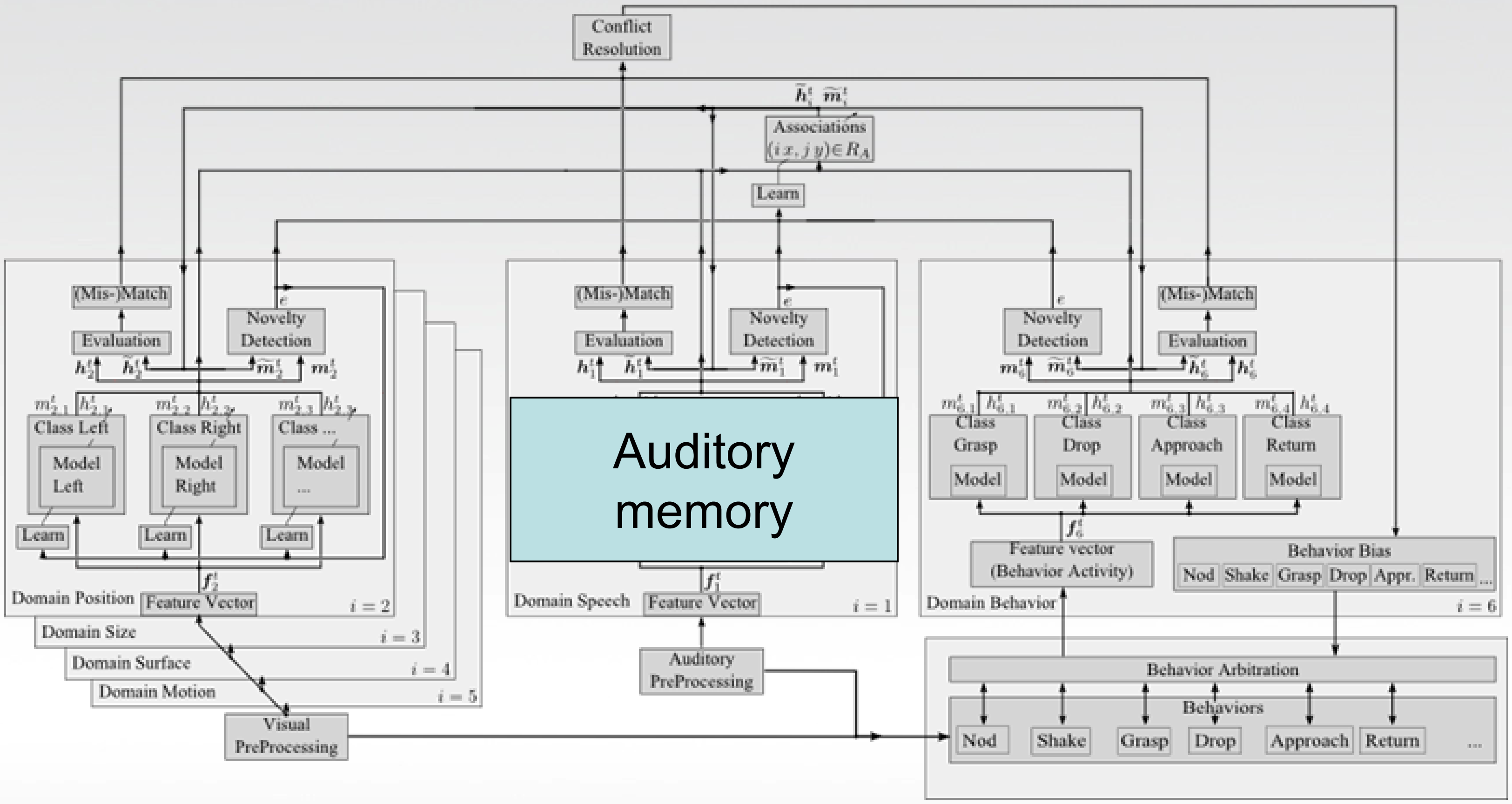
Schematics of Abstraction Layer



Major Architecture Elements

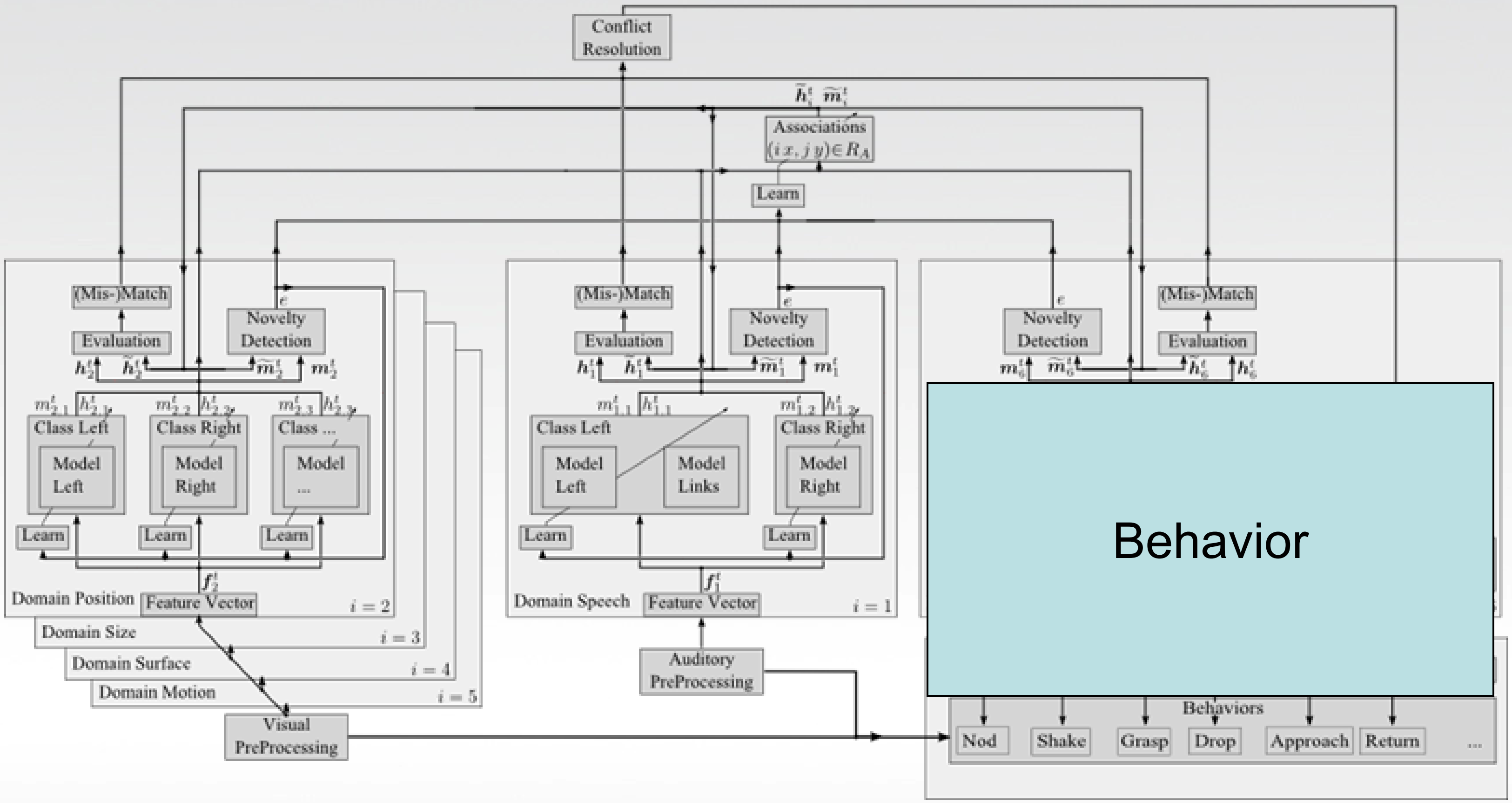


- Vision Memory: Proto-objects w/ size, position, orientation, …



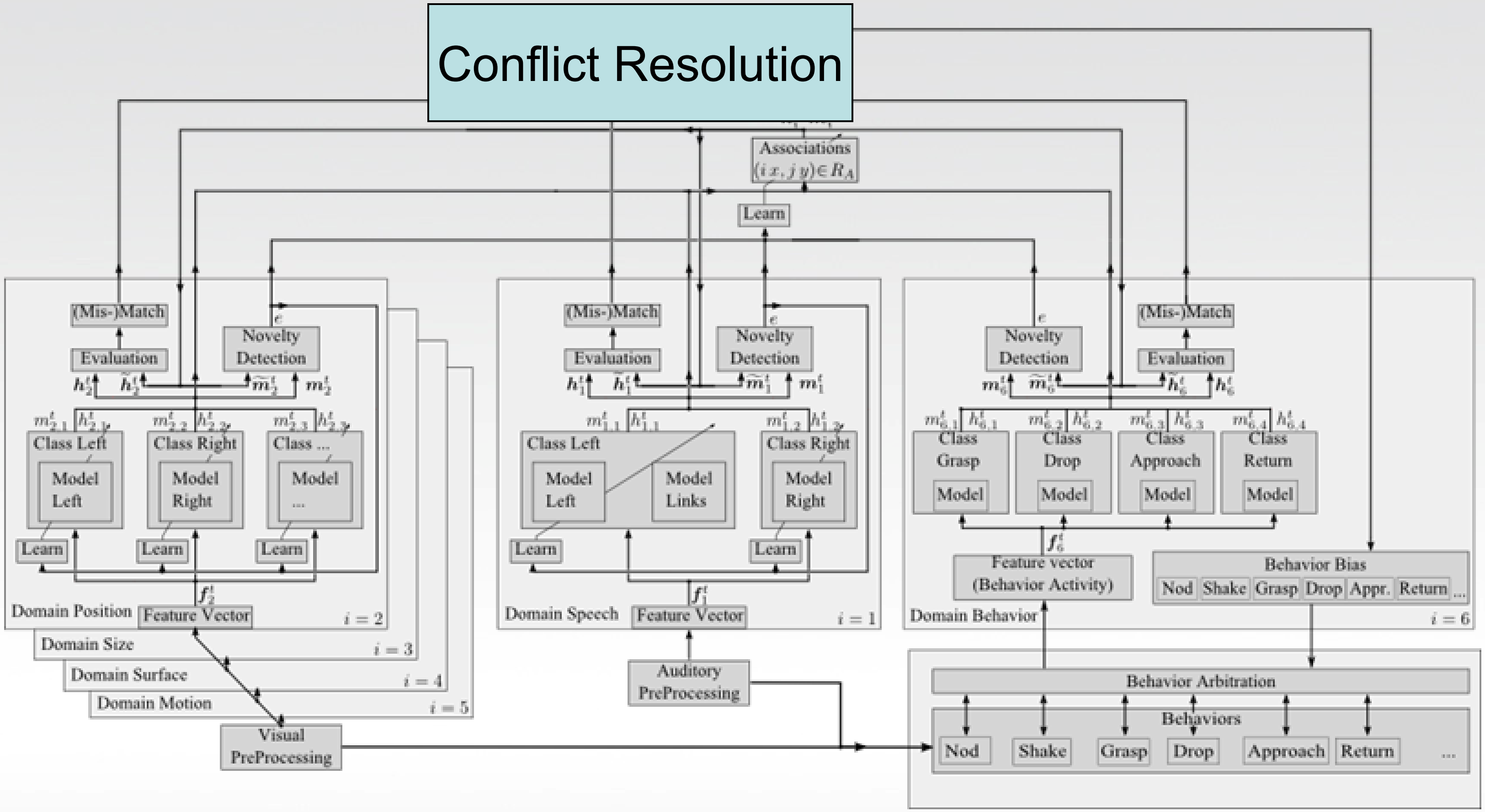
- Auditory Memory: Word-level Hidden Markov Models

Major Architecture Elements



- Behavior: Competitive activation dynamics with top-down bias

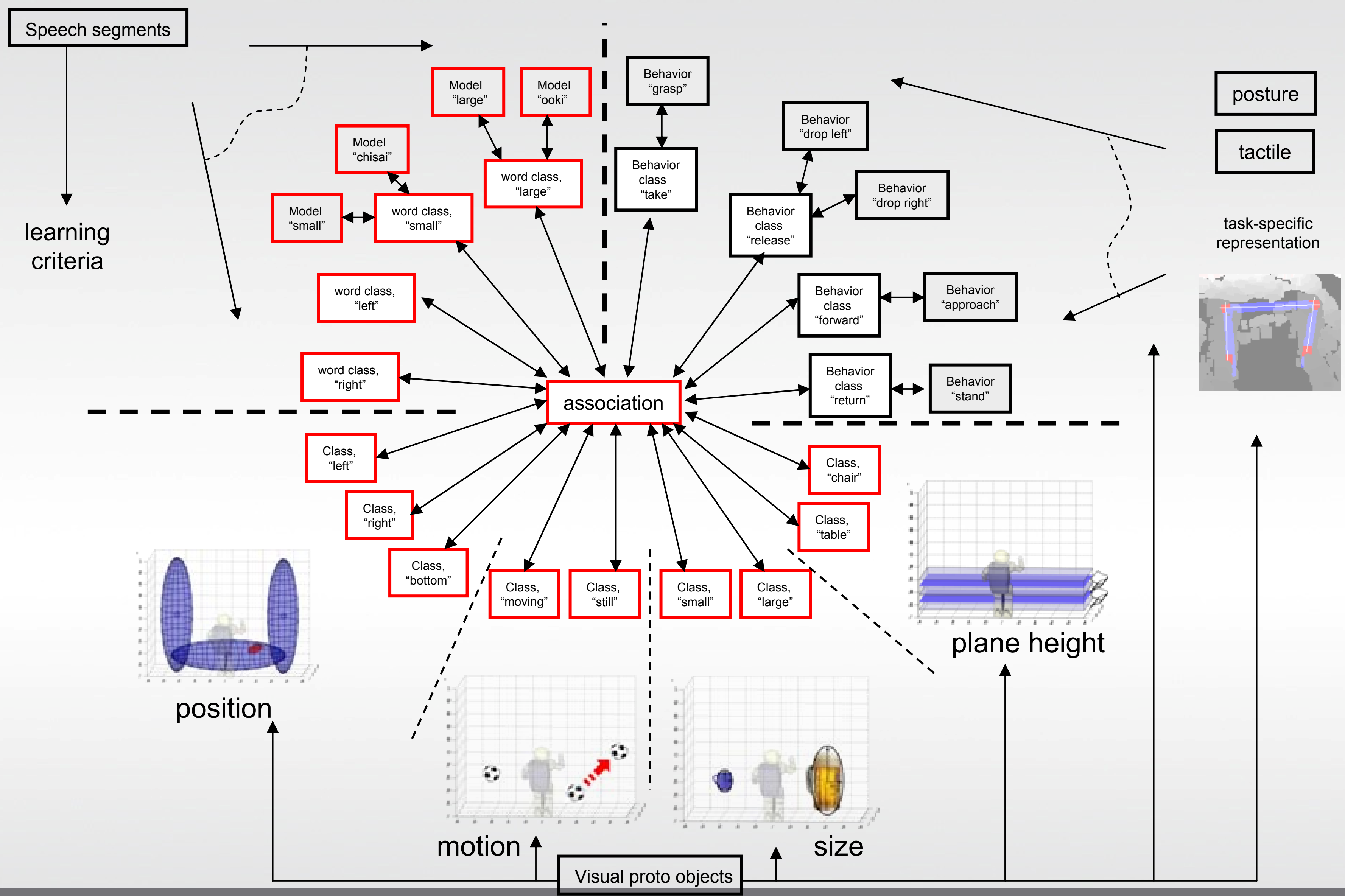
Major Architecture Elements

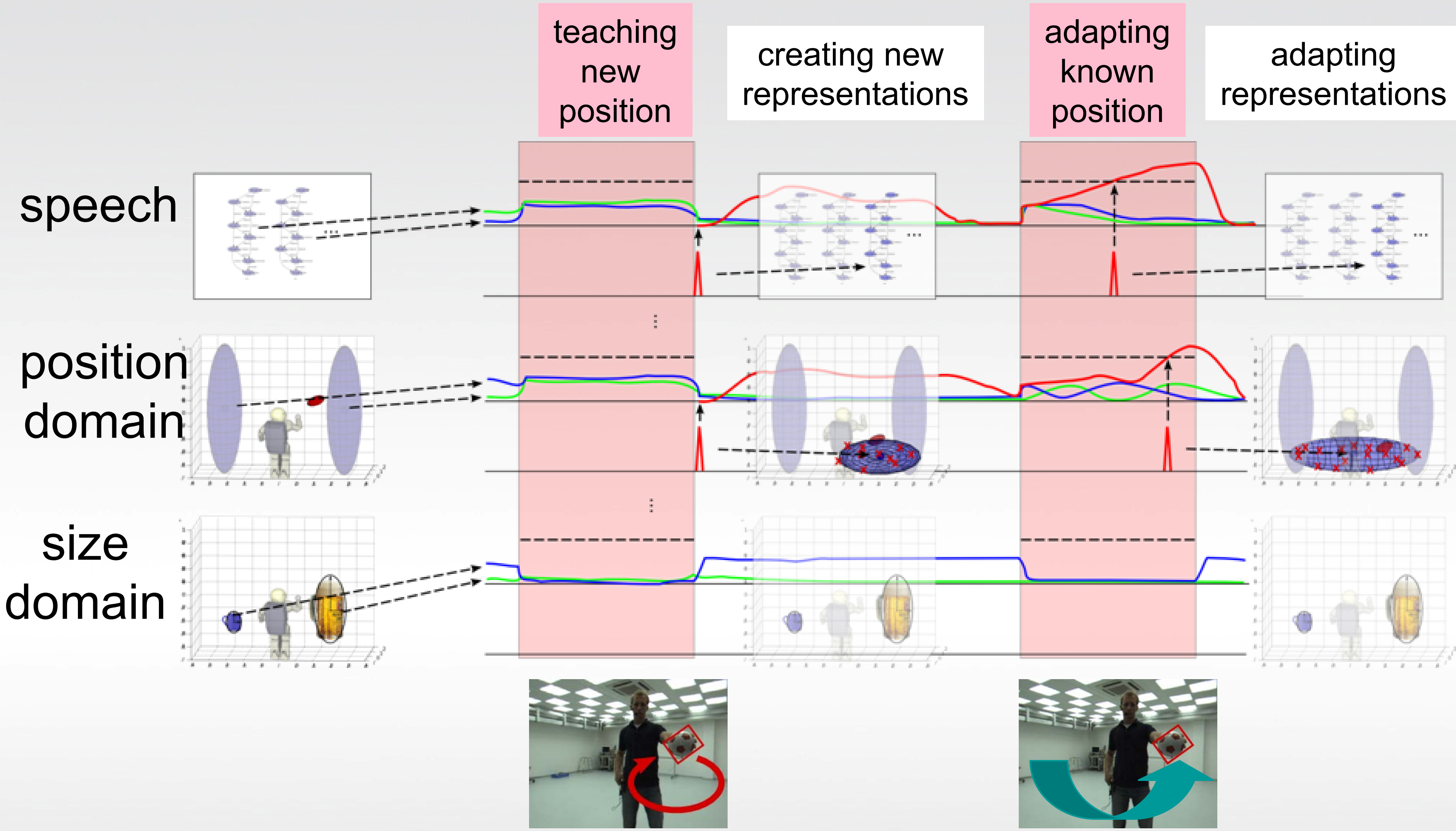


- Conflict Resolution: Trigger global resolution of previously determined mismatch

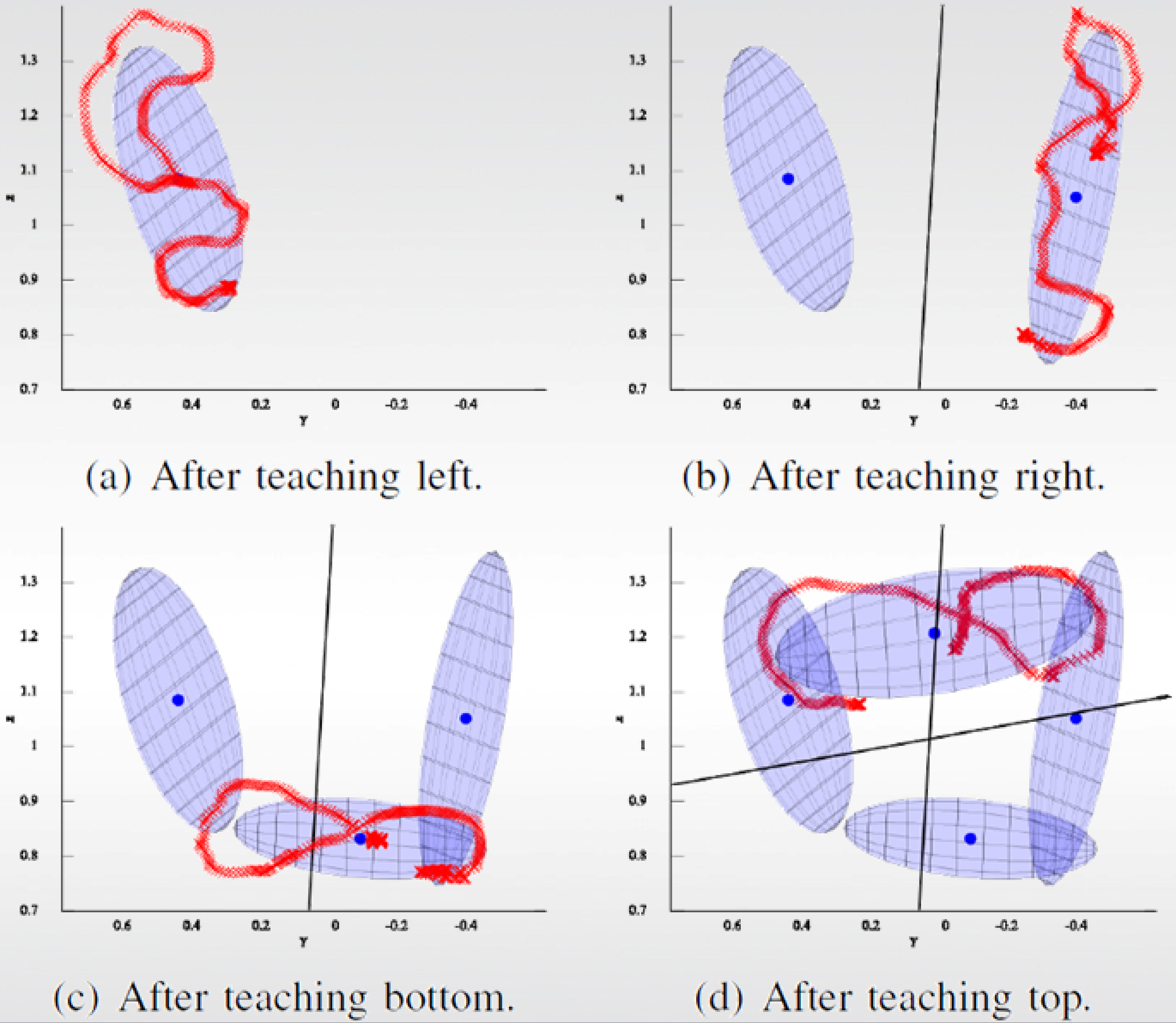
- All classes are homogeneously encoded like in the human cortex
- Population code are the common basis, i.e. linearly encoded activation vectors or matrices for each cue
- This forms the basis for cross-modal processing
- Classes can be any derived distinction of the features continuously delivered from the reactive layer

Representations Overview




teaching new position
creating new representations
adapting known position
adapting representations
speech
position domain
size domain

Overlapping Representations & Dependency Groups



(a) After teaching left.

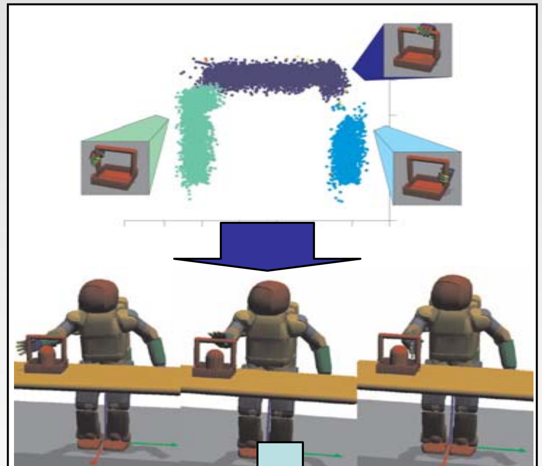
(b) After teaching right.

(c) After teaching bottom.

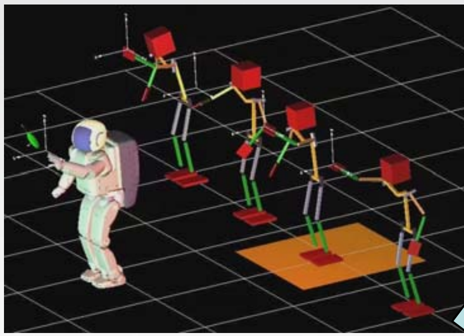
(d) After teaching top.

Grasping

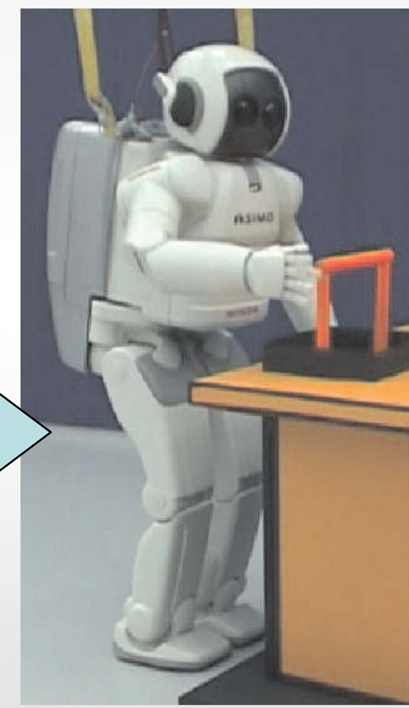
Learnt behavioral relevant object-body relations



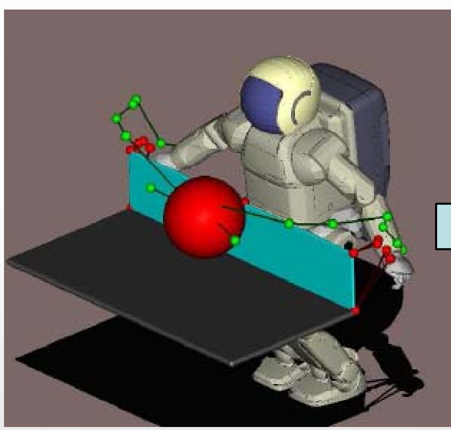
Internal Prediction & decision making

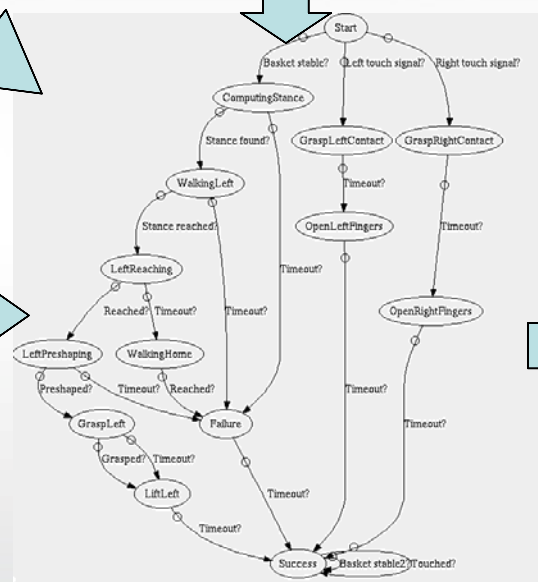


Reaching & grasping in dynamic environments



Planning optimal movements





Sequence of movement primitives

Results:

- Close talk microphone free interaction
- Learning of visual property classes, auditory labels and associations
- Steps into physical interaction

no grasp
1:26min

explanation
6:01min

grasp
2:23min

Heckmann, Brandl, Schmüdderich, Domont, Bolder, Mikhailova, Janssen, Gienger, Bendig, Rodemann, Dunn, Joublin, Goerick: *Teaching a humanoid robot: head-set free speech interaction for audio-visual association learning,* RO-MAN 2009

Goerick, Schmüdderich, Bolder, Janßen, Gienger, Bendig, Heckmann, Rodemann, Brandl, Domont, Mikhailova: *Interactive Online Multimodal Association for Internal Concept Building in Humanoids,* Humanoids 2009

Summary & Outlook

- What you have seen
  - One focus of ALIS 3
  - Teaching and evaluation “large” concept
  - Teaching and evaluation of “take” and “release” concept
  - Relative positions (“left”, “right”, “top”, “bottom”)
  - Learning of synonyms (ooki, chisai, groß, klein)
  - Interaction without close talk microphone
- What you have not seen
  - Motion state (“moving”, “still”)
  - Height of supportive plane (“table”, ”chair”)
- Open issues
  - More elaborate internal representations
  - Active hypothesis testing for resolution of ambiguities
  - Acquisition of behaviors
  - Intrinsic Motivations for What and When to learn

Thank you for your attention