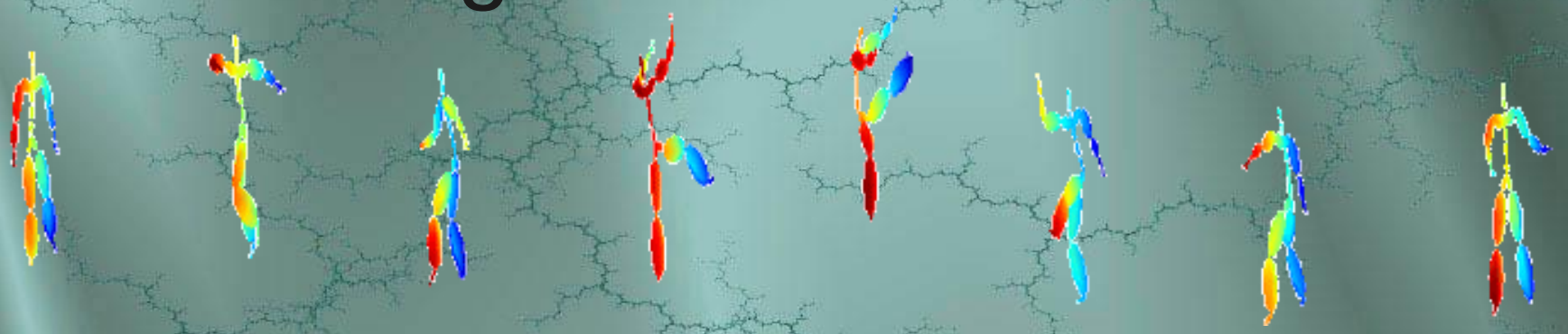
Human Activity Language: Grounding Concepts with a Linguistic Framework



Gutemberg Guerra-Filho

{guerra@cs.umd.edu}

Yiannis Aloimonos

{yiannis@cfar.umd.edu}

Computer Vision Laboratory

Department of Computer Science

University of Maryland

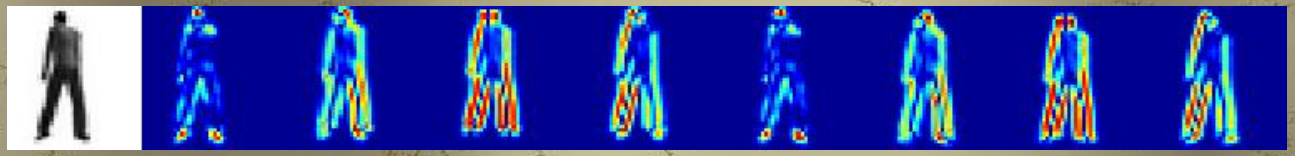
Theme

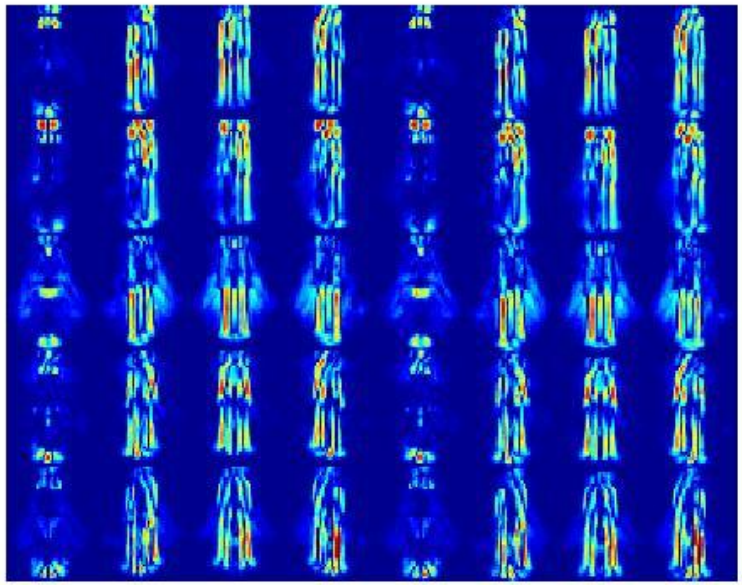
- To close the semantic gap in multimedia technologies, we need to **understand human action**
- There are at least 3 spaces devoted to human action: The Visual, the Motoric, and the Language Space.
- Each of these spaces is characterized by a distinct language, with its own alphabet, words, and syntax.

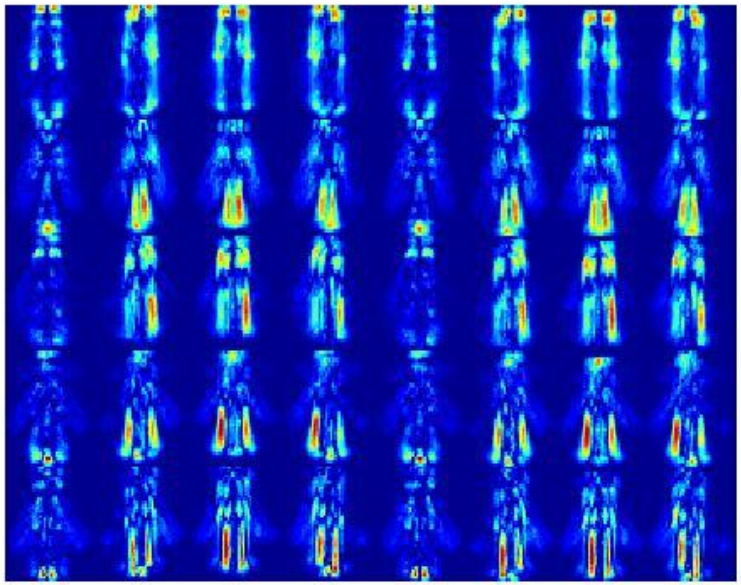
Initial Meeting

- Multimedia
- Semantics
- Semantics arises from human action
- Big brother problem

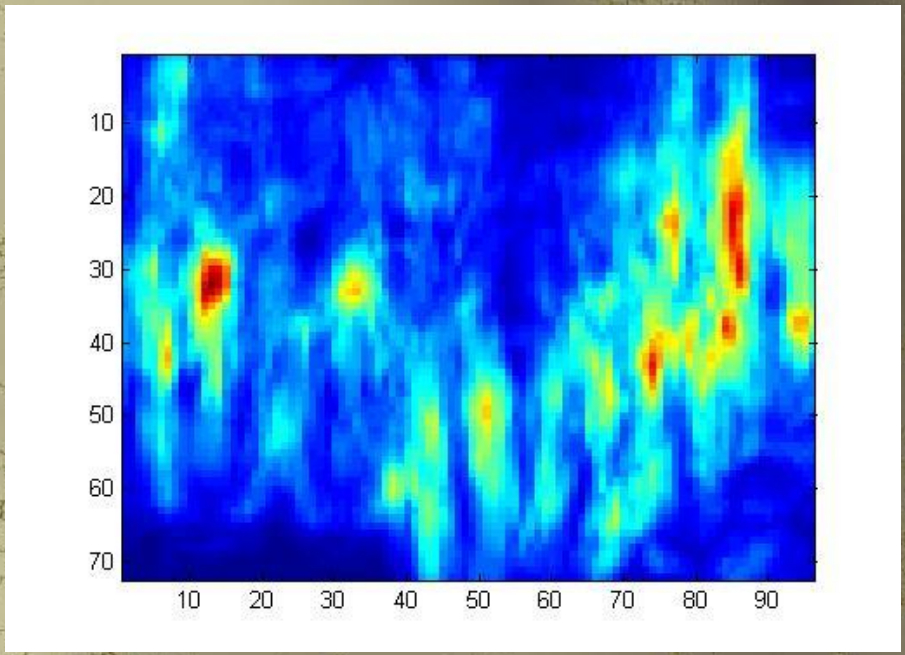
VHF's: Visual Human Filters





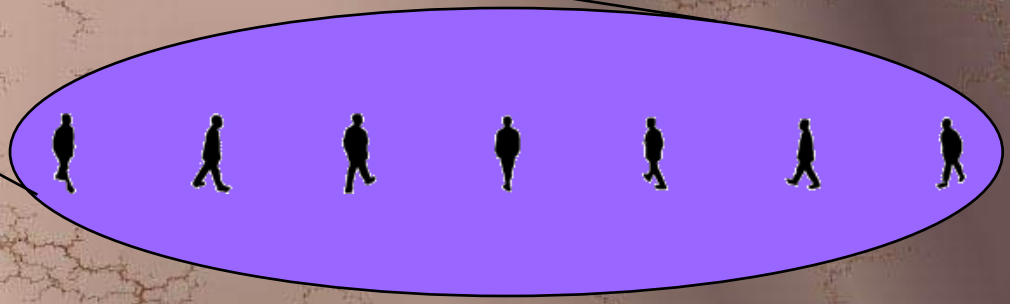


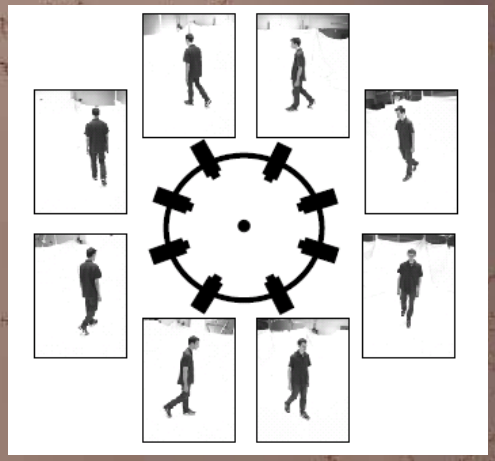
Applying the VHF's



Visual Approach: Sequences of poses

p_1 Stand	
p_2 Bent Knees	
p_3 Legs Apart(1)	
p_4 Legs Together	
p_5 Legs Apart(2)	
p_6 Kick Leg Behind	
p_7 Kick Leg Front	
p_8 Kick Legs Together	
p_9 Kneel	
p_{10} Half Squat Down	
p_{11} Squat	
p_{12} Half Squat Up	
p_{13} Half Bend Down	
p_{14} Full Bend	
p_{15} Half Bend Up	
p_{16} Start Sit Down	
p_{17} Half Sit Back	
p_{18} Full Sit	
p_{19} Half Sit Front	
p_{20} Start Sit Up	





What are “Key” poses?

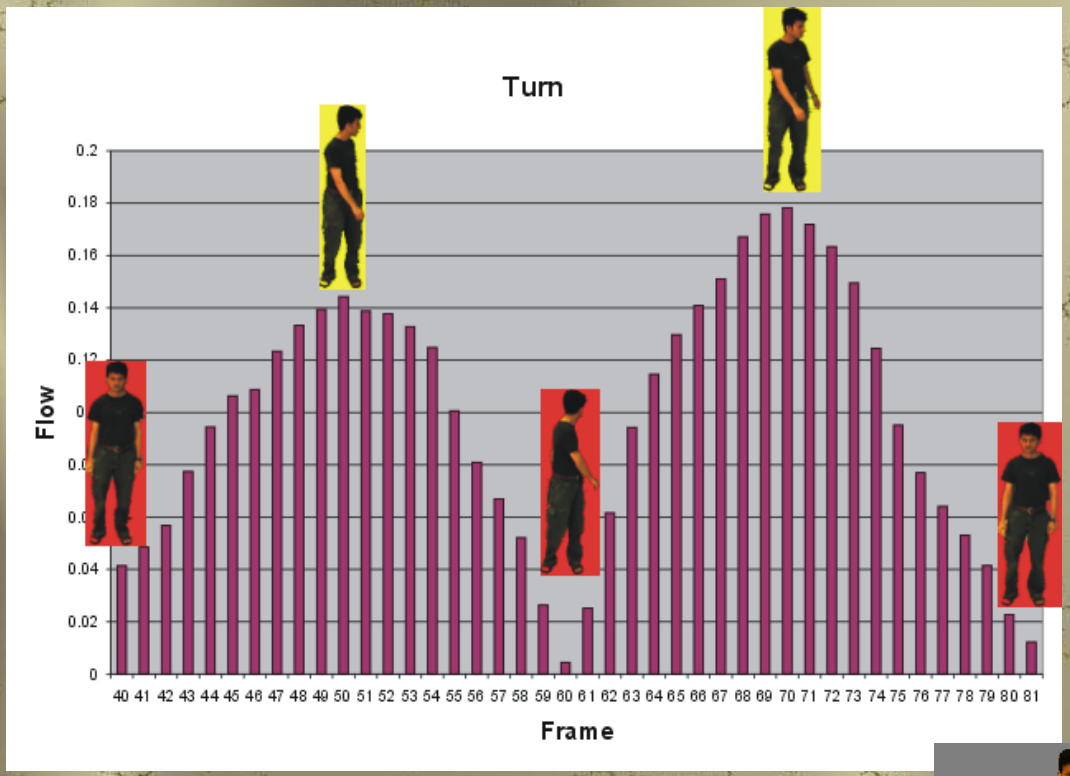
- Extremal poses of the body.
- How are they found?
- Single-view example:

Video

Horizontal motion
Red = right
Blue = left

Vertical motion
Red = Down
Blue = Up

What are “Key” poses?



Mean flow magnitude (in person’s reference frame)

Key frames

Pose Grammar

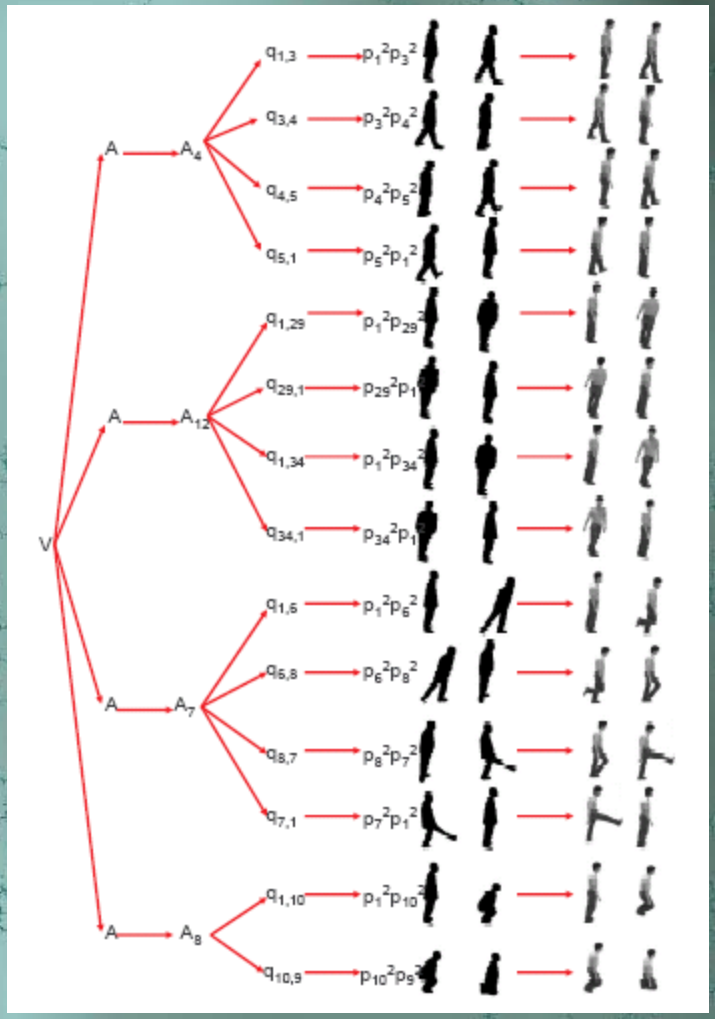
- Probabilistic context-free Grammar (PCFG).

$Start \rightarrow V$	$p = 1$
$V \rightarrow VA \mid A$	$p = 1/2$
$A \rightarrow A_1 \mid A_2 \mid .... \mid A_g$	$\forall i, p(A_i \mid A) = 1/g$
$A_i \rightarrow q_{ab} q_{bc} q_{cd} ...$	$p(q_{ab} q_{bc} q_{cd} ... \mid A_i) = 1$
$q_{cd} \rightarrow p_c^u p_d^v$	$\sum_{\substack{allowed \\ u,v}} p(p_c^u p_d^v \mid q_{cd}) = 1$
$p_i^v \rightarrow s_k$	$p(s_k \mid p_i^v)$ *obtained at runtime*

Rules created from training data

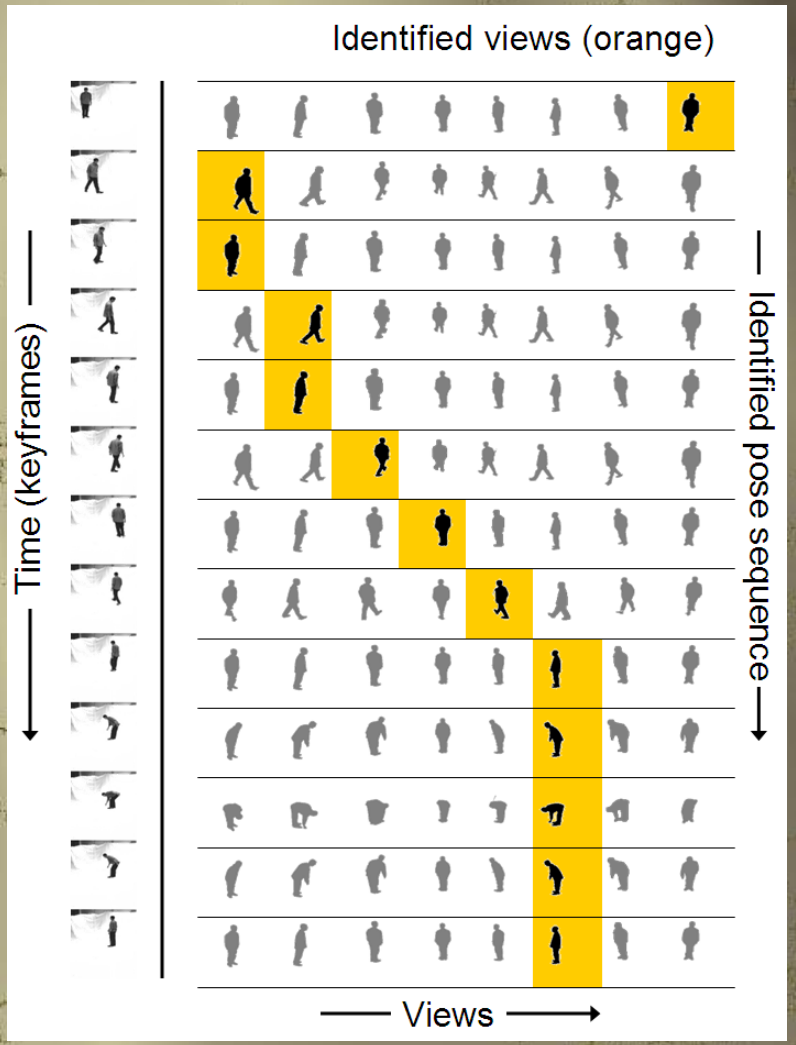
Parse an input video

1. Key frame detection.
2. Silhouette matching on keyframes.
3. Computation of $P(s_k|p_i^v)$ as shown earlier.
4. Probabilistic parsing using the PCFG.



Detect view changes

Person walks along a circle



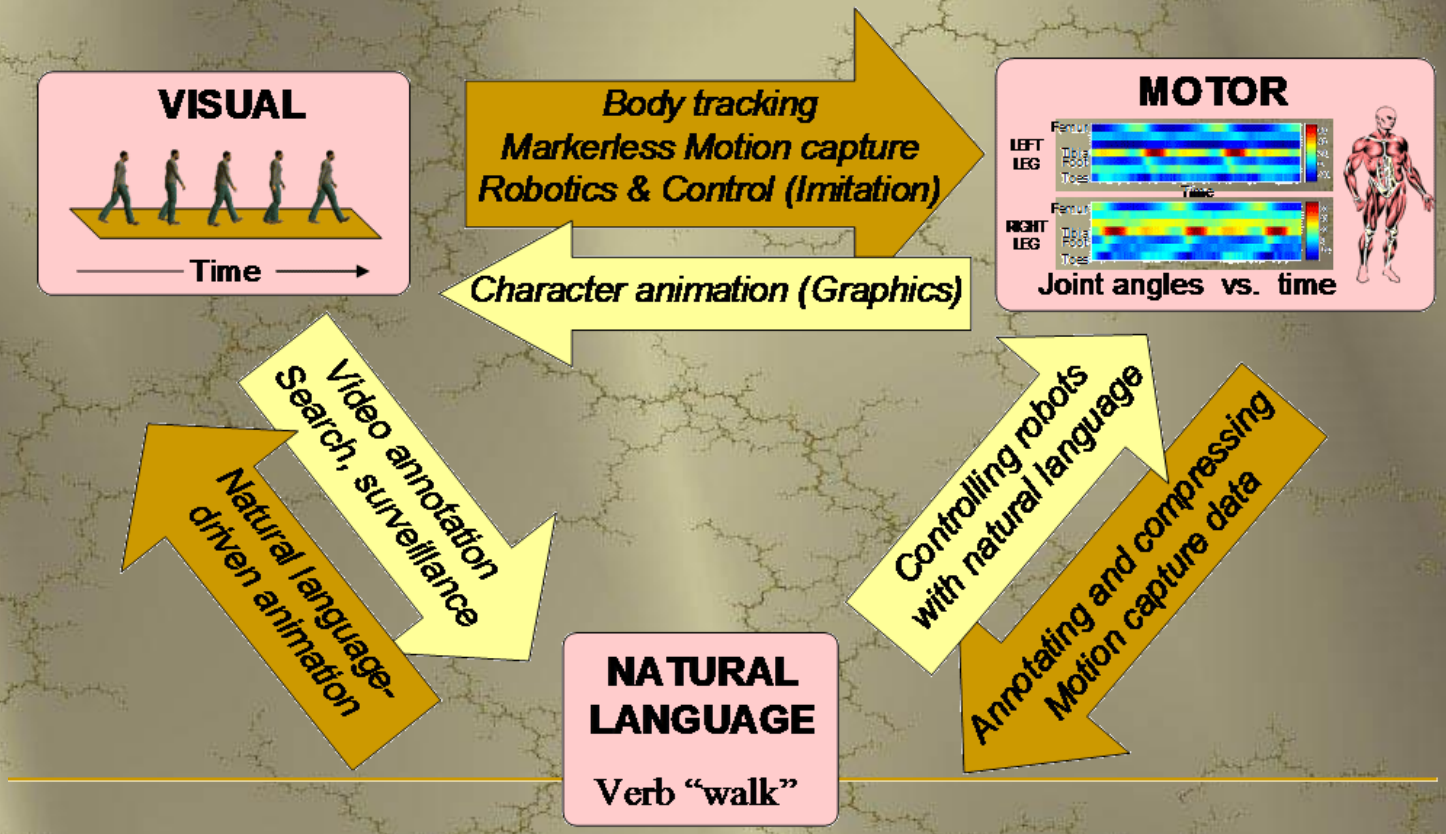
One year later…

- That's What I Found

- Easier to solve the visual action problem by going first through the motor action problem.

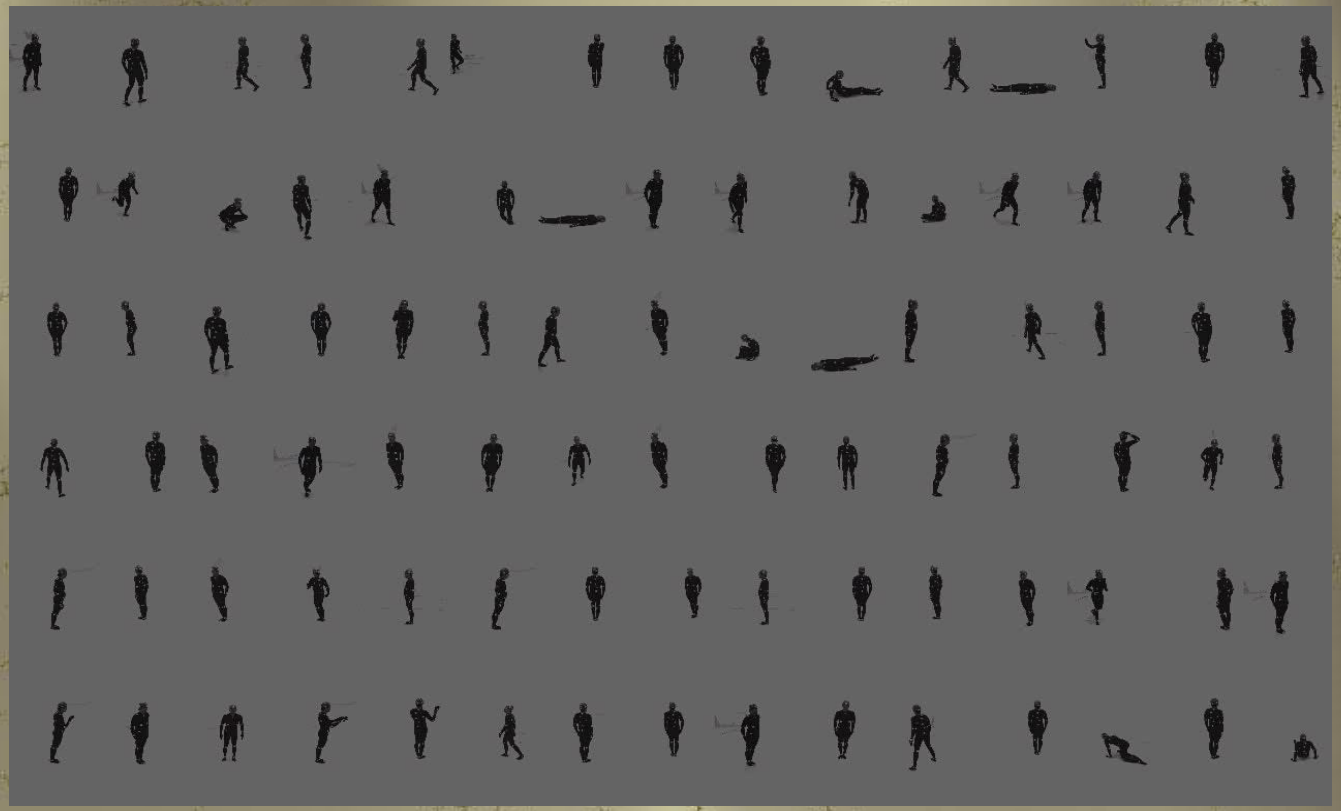
- Human Activity Language (HAL): a new language for human activity.

Spaces for Human Action

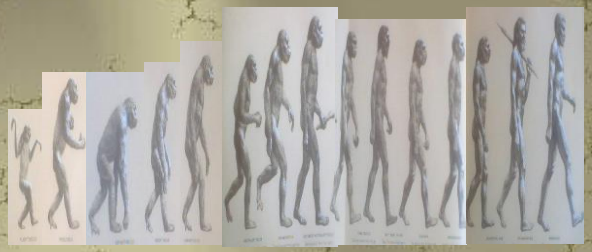


Hyperempiricism in Computer Science

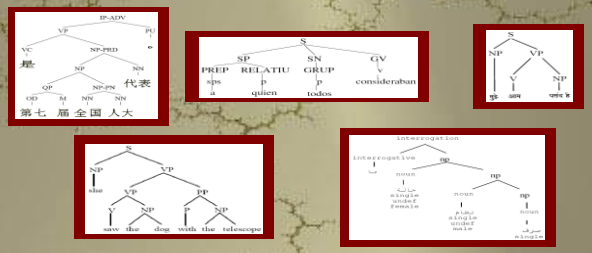
Problem



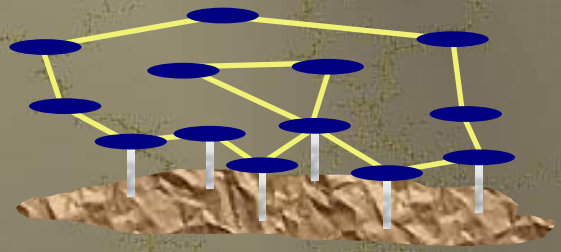
Language Origin



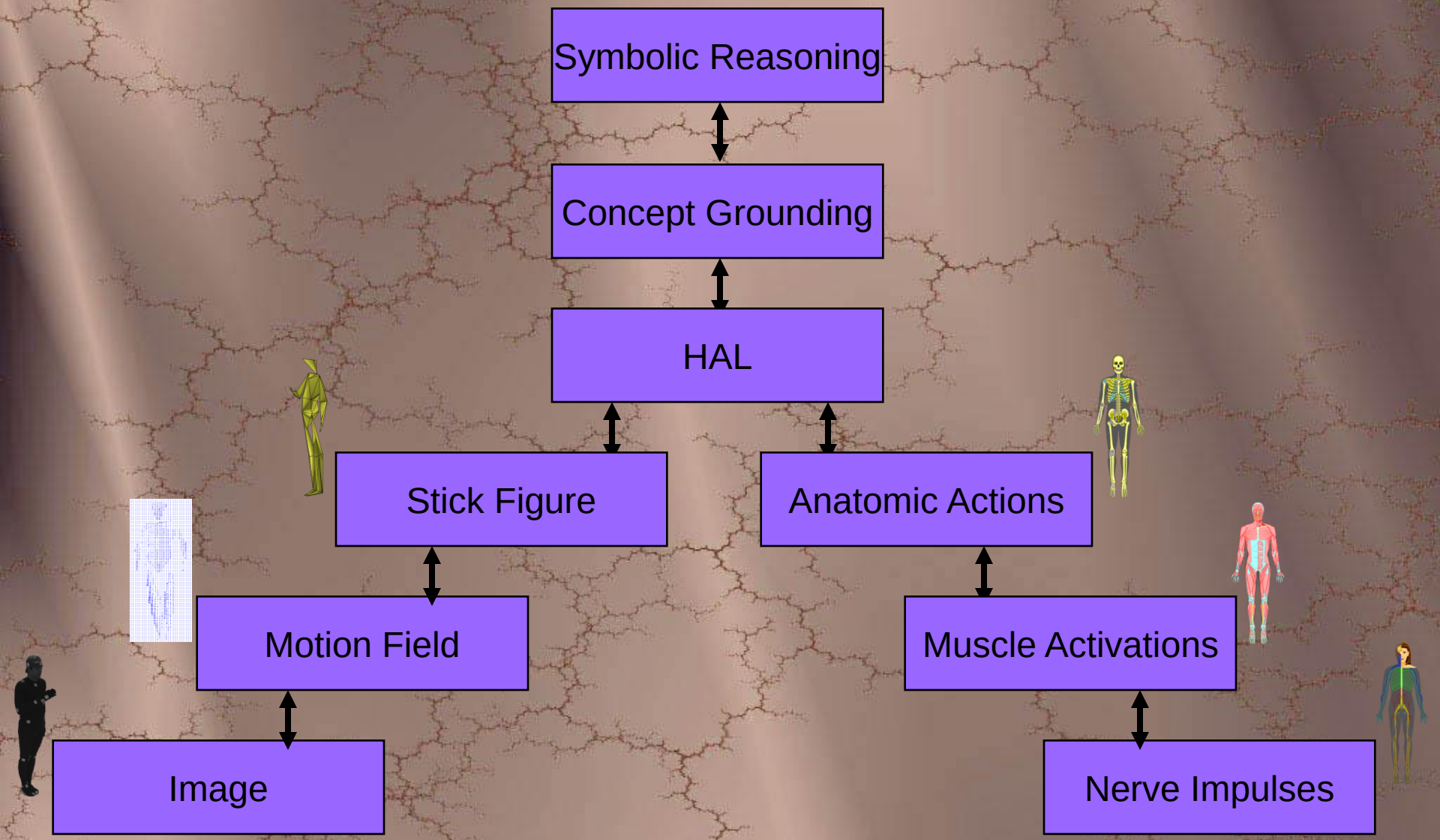
Universal Grammar



Concept Grounding

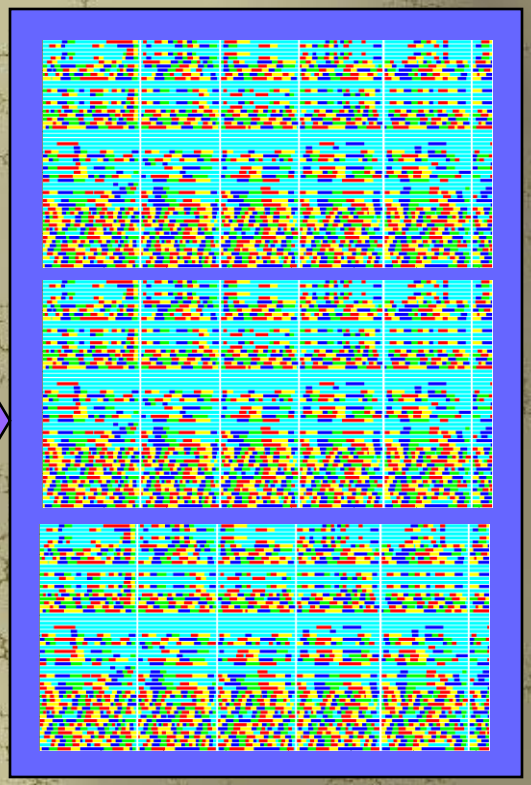


Sensory-Motor Intelligence

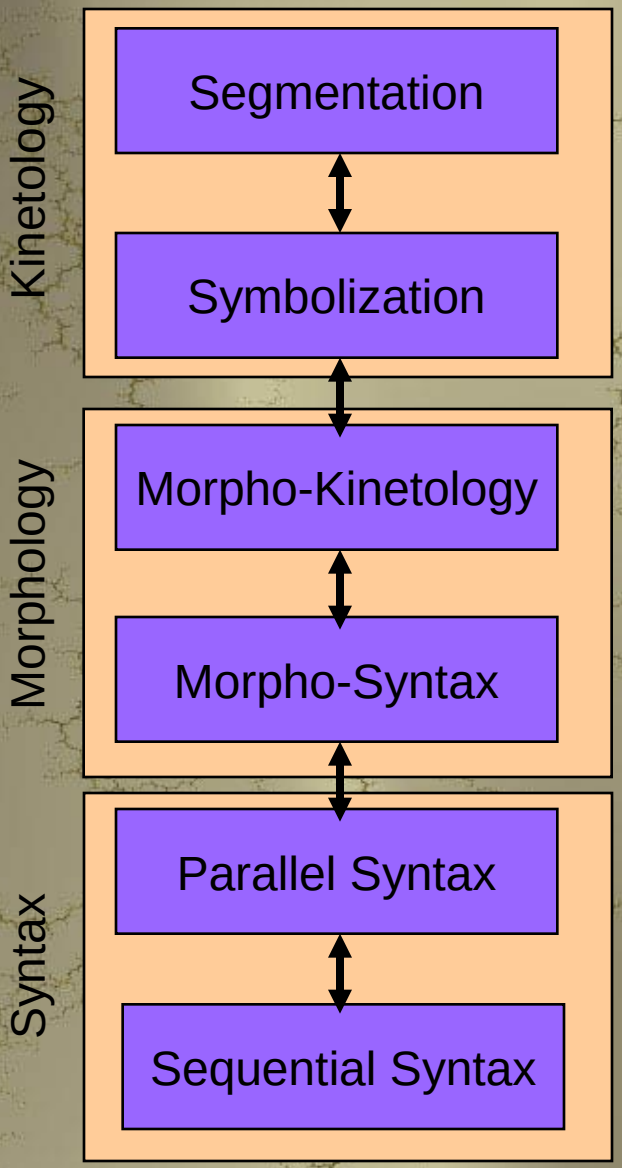


Praxicon

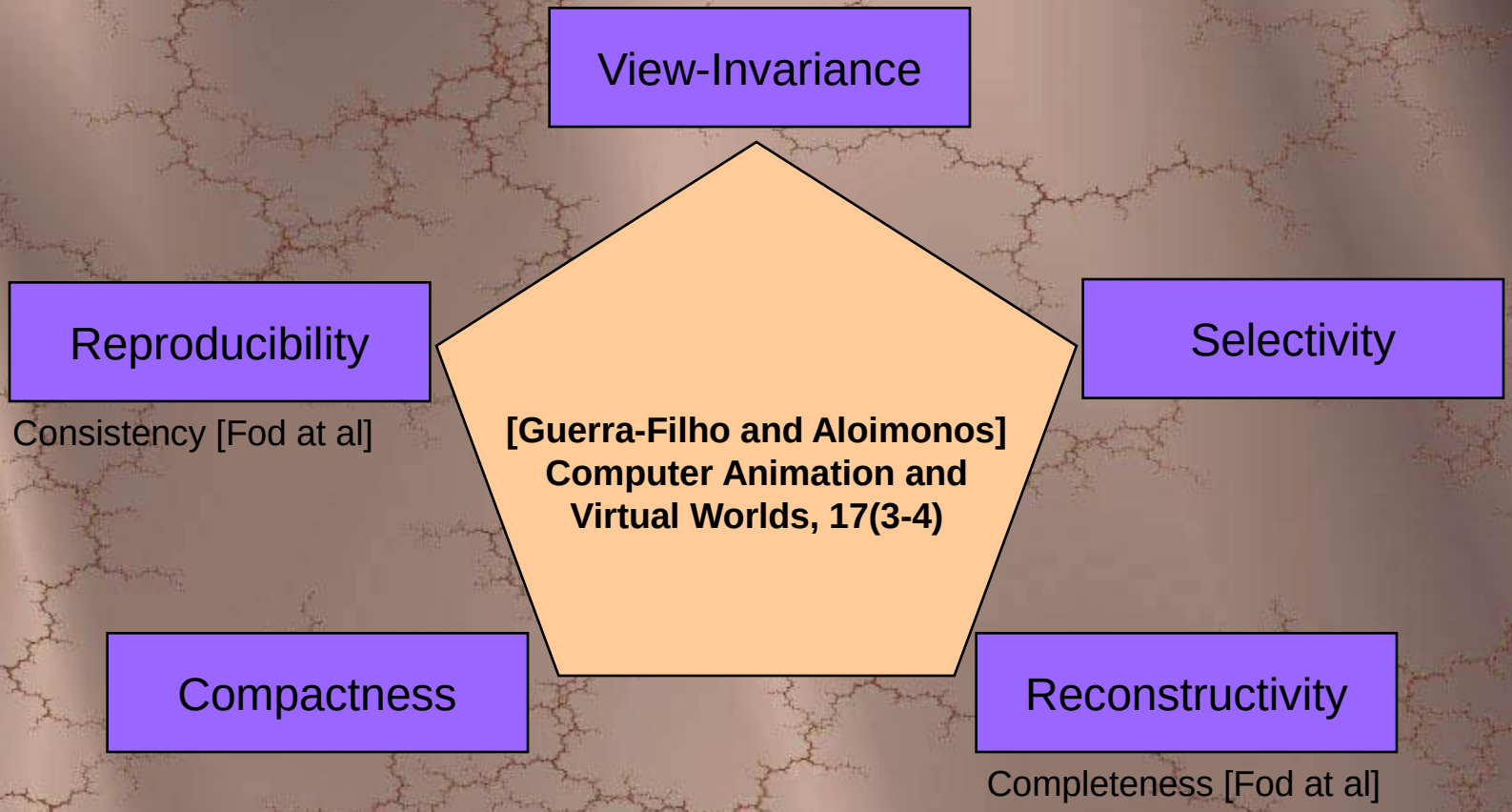

head & neck capture
shoulder capture
arm, elbow & wrist capture
back & pelvis capture
hand & finger capture
leg & knee capture
ankle, foot & toe capture



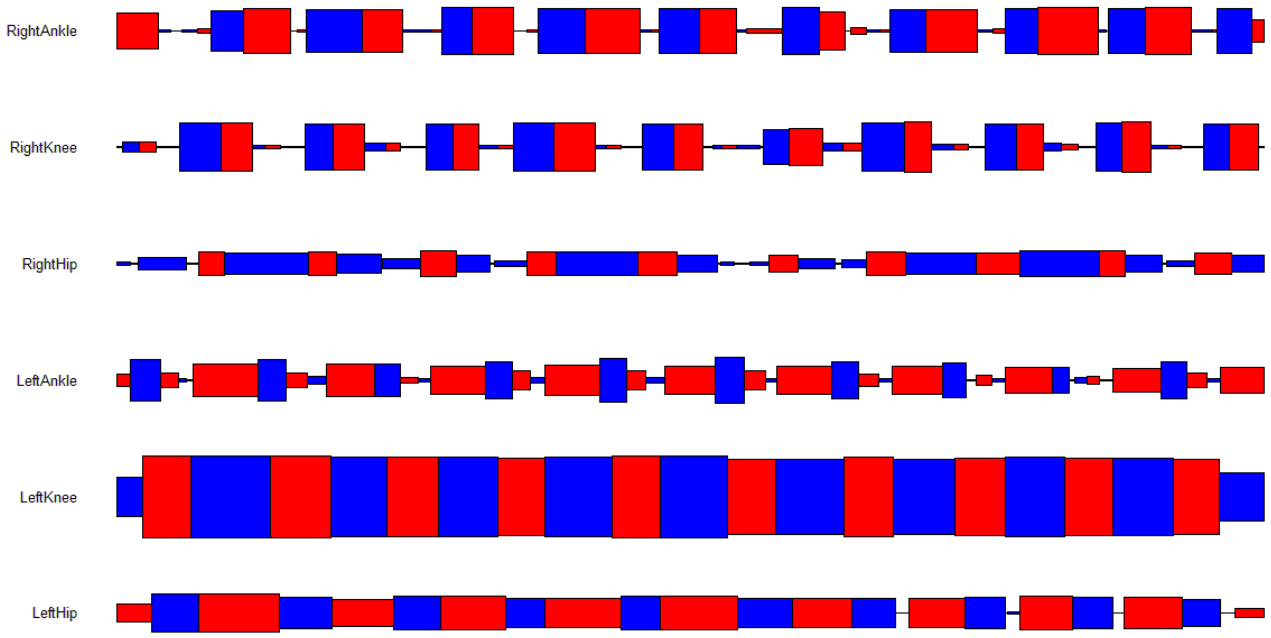
Human Activity Language



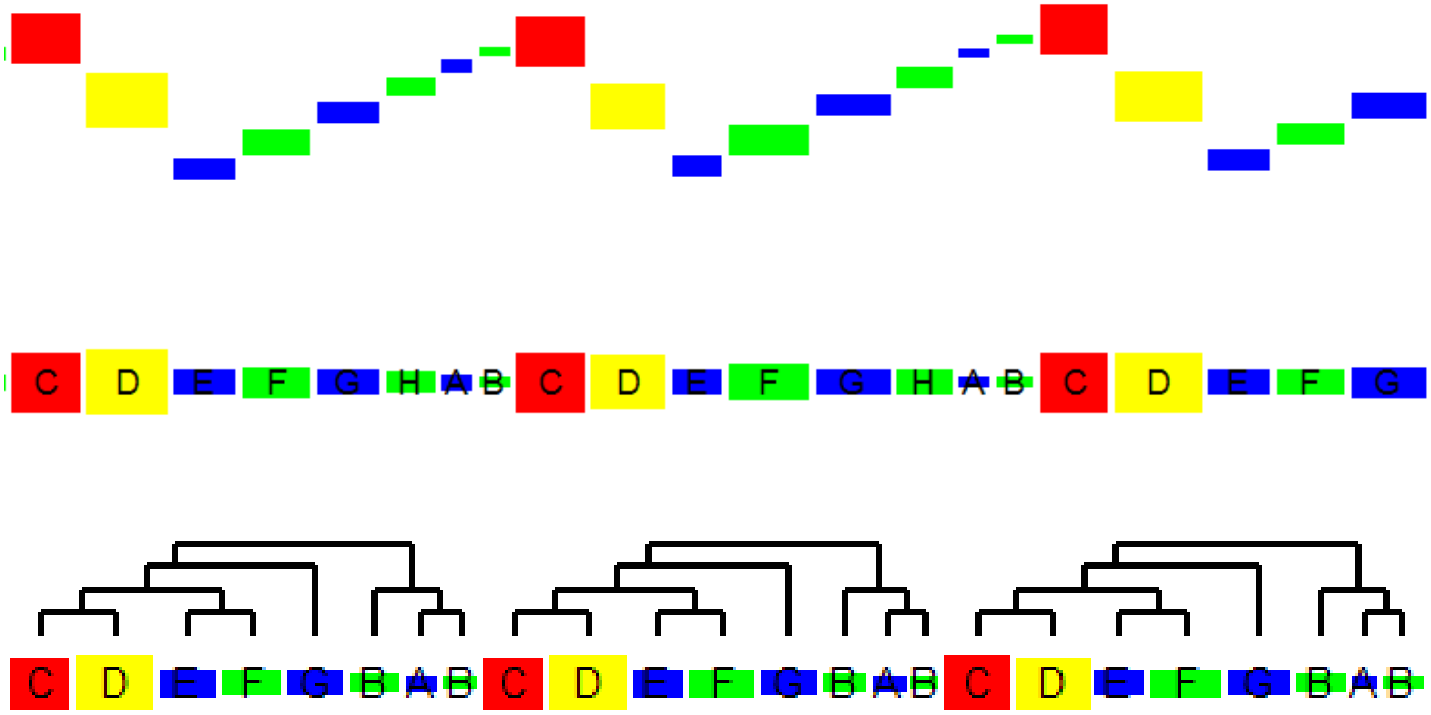
Kinetology


View-Invariance
Reproducibility
Consistency [Fod at al]
[Guerra-Filho and Aloimonos]
Computer Animation and
Virtual Worlds, 17(3-4)
Selectivity
Compactness
Reconstructivity
Completeness [Fod at al]

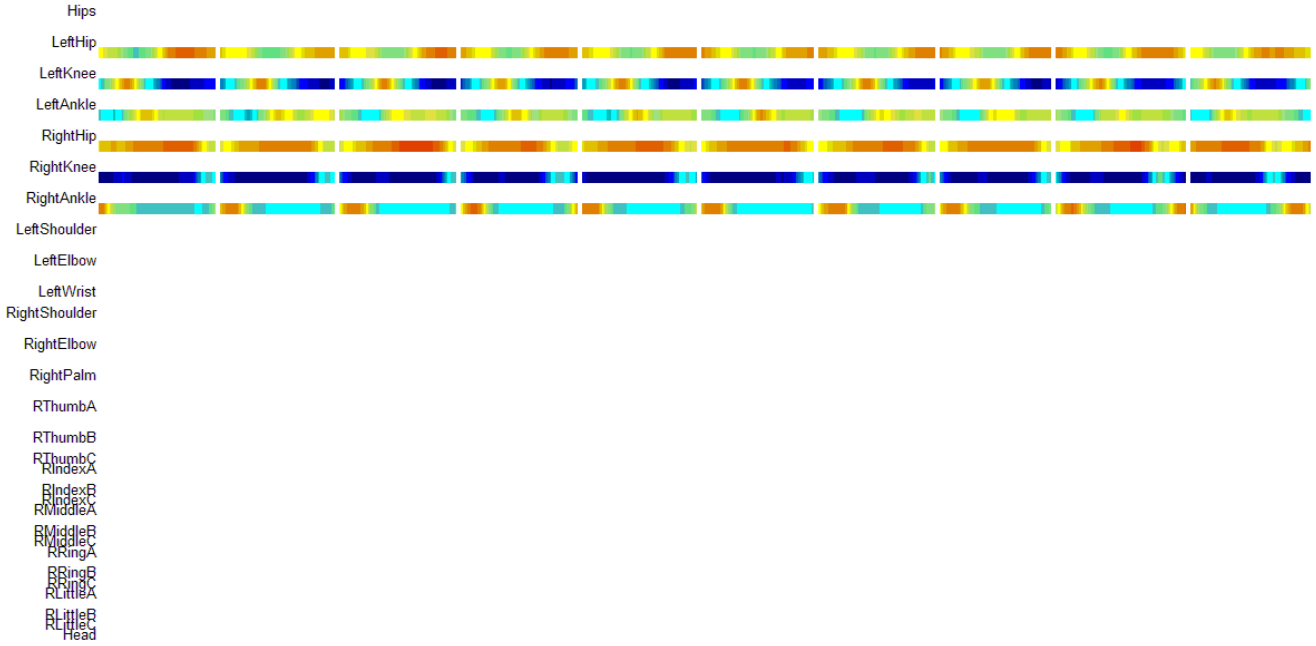
Segmentation



Symbolization

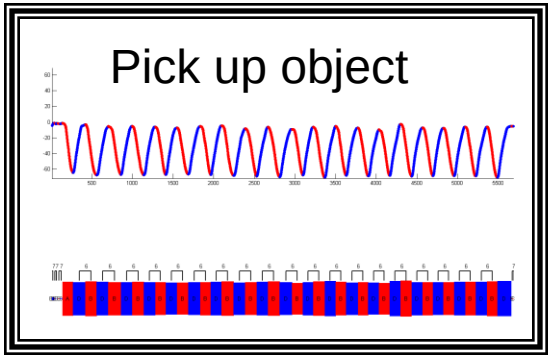


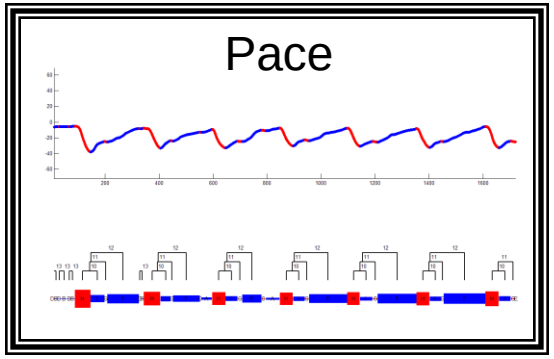
Morphology

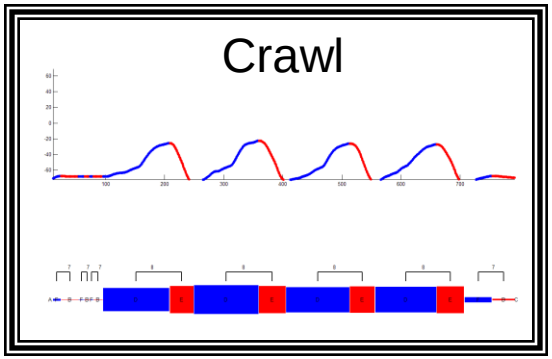

Hips
LeftHip
LeftKnee
LeftAnkle
RightHip
RightKnee
RightAnkle
LeftShoulder
LeftElbow
LeftWrist
RightShoulder
RightElbow
RightPalm
RThumbA
RThumbB
RThumbC
RIndexA
RIndexB
RIndexC
RMiddleA
RMiddleB
RMiddleC
RRingA
RRingB
RRingC
RLittleA
RLittleB
RLittleC
Head

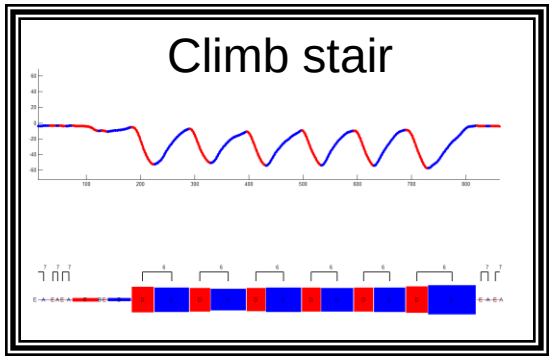
Morpho-kinetology

Right Hip Flexion-Extension

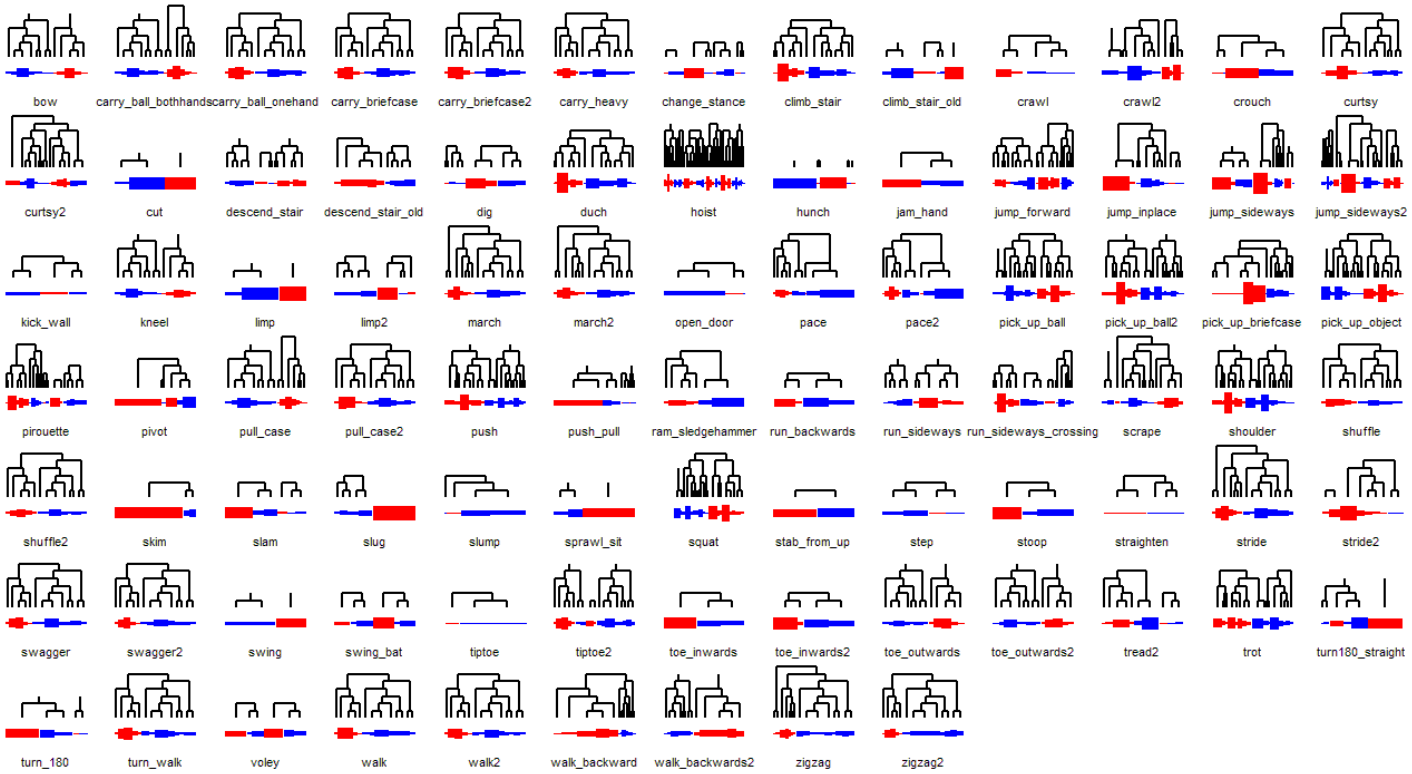




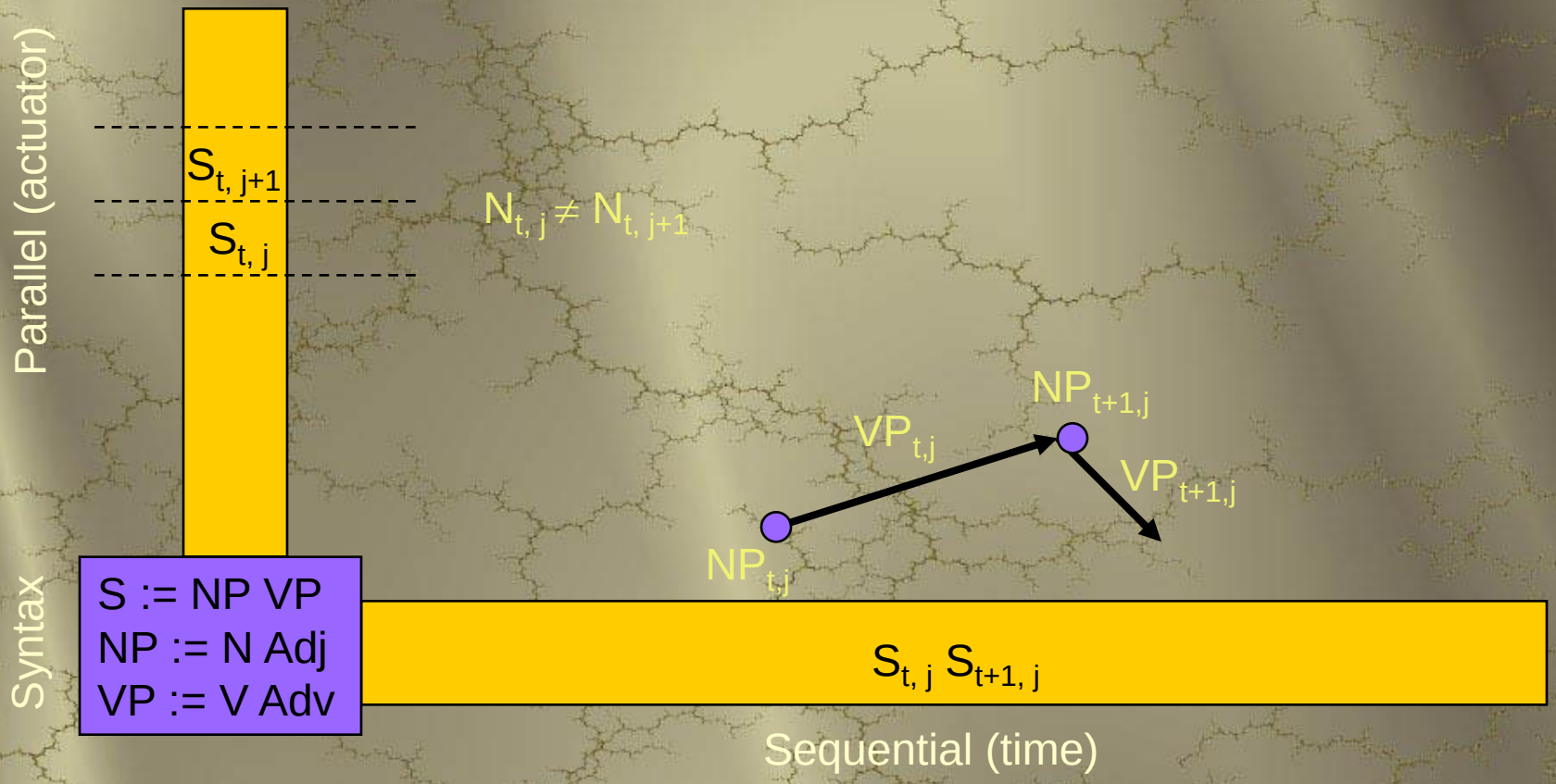




Morpho-syntax



Syntax

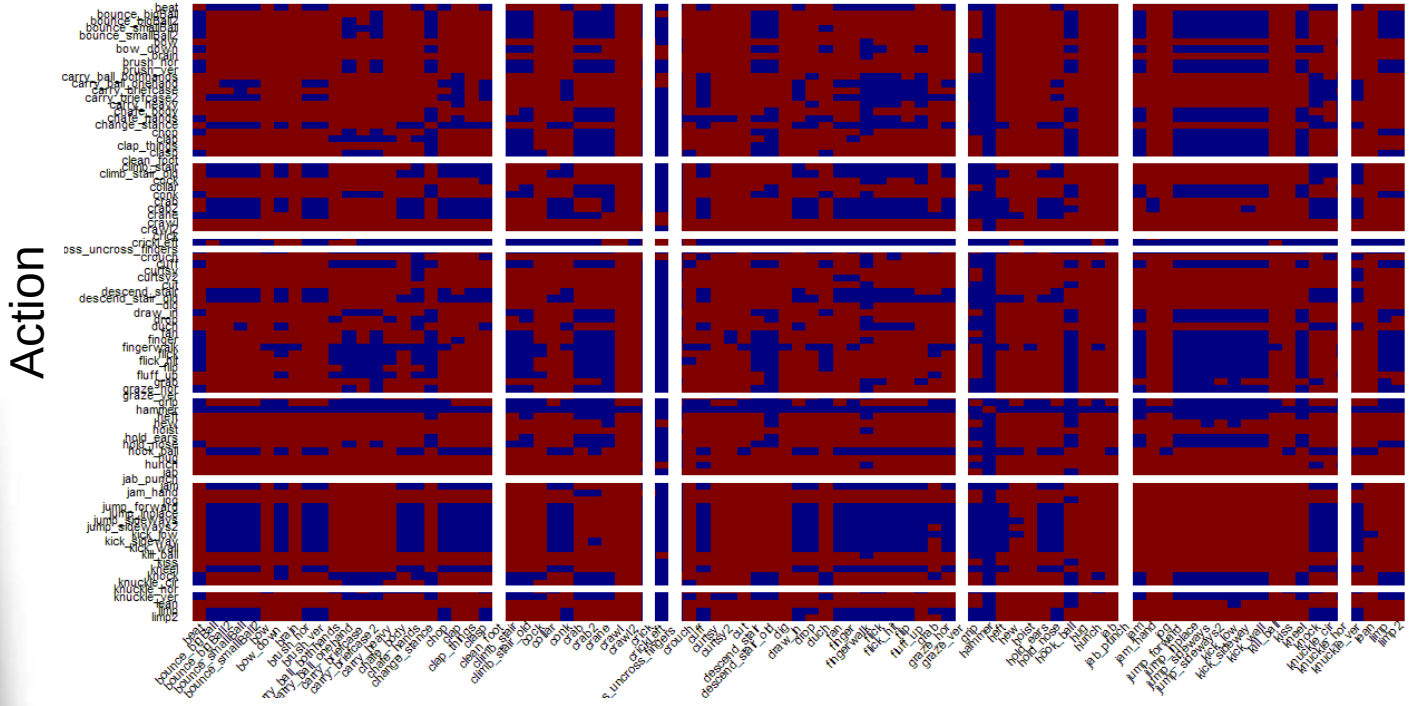


- Noun: Body parts active during the execution of a human activity
- Verb: Changes each active joint experiences during the activity execution
- Adjective: Specifies the initial state of the active joints (initial posture)
- Adverb: Modifies verb with purpose of generalization

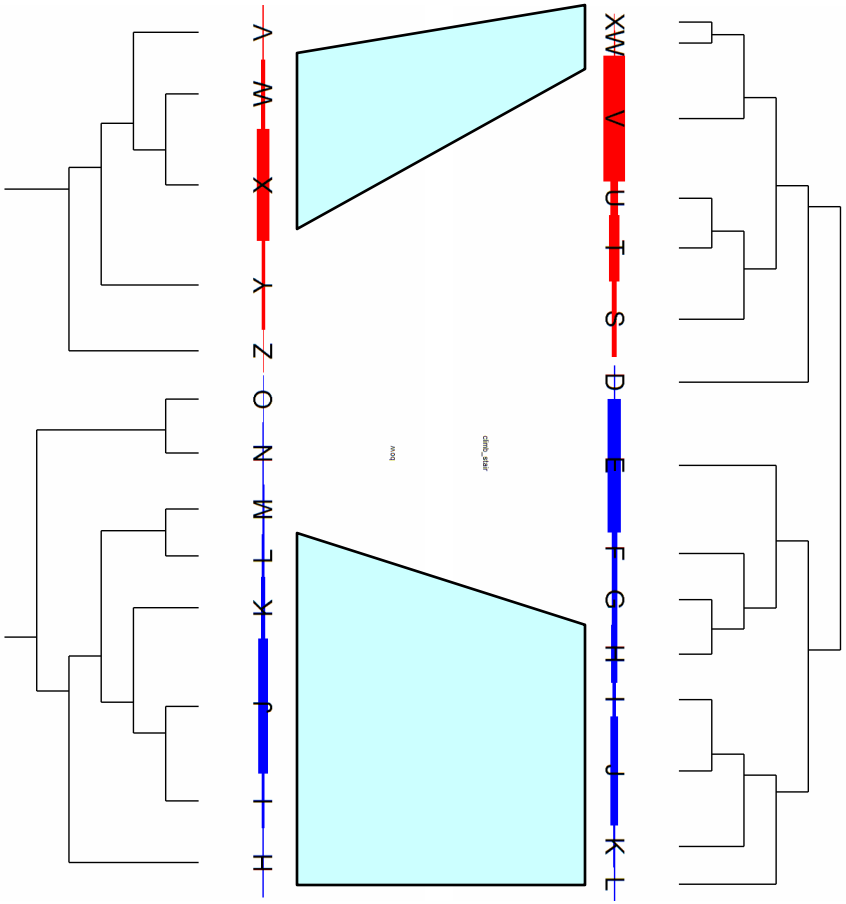
Parallel Syntax

{crick, cross fingers, knuckle, graze, jab, clean foot}

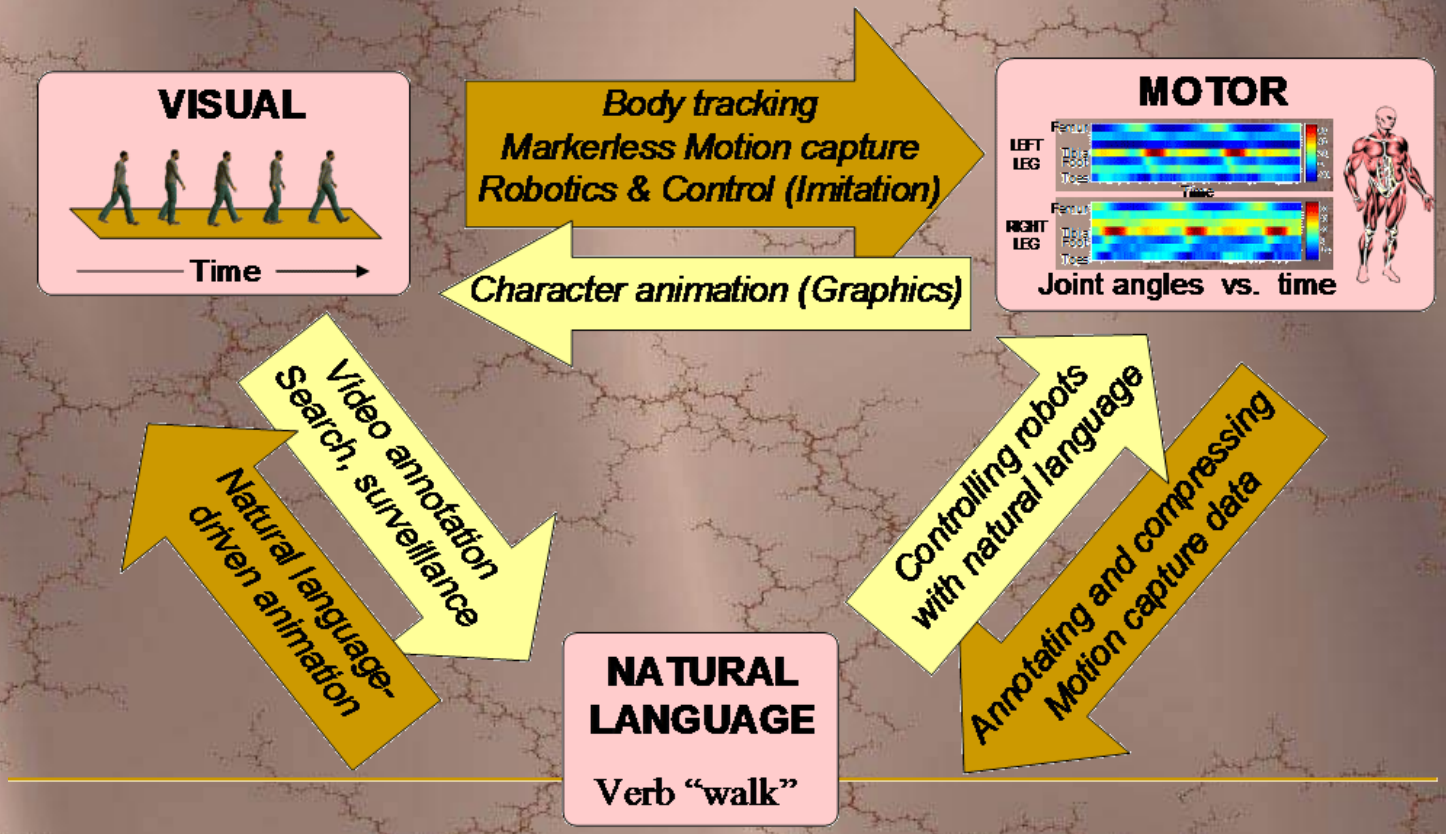
Constraint Matrix



Sequential Syntax



Conclusions



Sensory-Motor Theories vs Symbolic Theories

The Behaviorome Project