

# Notes on Semantic Hierarchical Temporal Memory for Perceptual, Motoric and Intentional Intelligence

*Ben Goertzel<sup>1,2</sup> and Shuo Chen<sup>2</sup>*

<sup>1</sup> Novamente LLC, Rockville MD, USA

<sup>2</sup> Cognitive Science Dept., Xiamen University, Xiamen, China

## **Abstract**

The Hierarchical Temporal Memory (HTM) approach to perception processing, pursued by prior authors including Jeff Hawkins and Itamar Arel, is expanded to encompass more abstract semantic forms of HTM, and to include separate, coupled hierarchies for perception, motor control and goal management. No experimental evidence or detailed calculations are presented; this is merely a conceptual paper describing some potentially promising ideas.

Firstly, an extension of the HTM approach to perceptual AI is suggested, in which a traditional "perceptual" HTM is coupled with a more abstract "semantic-perceptual HTM", while retaining the same essential spatiotemporal pyramidal architecture. This extension is argued to provide more flexible visual pattern recognition, as well as easier interlinkage between HTMs and semantic networks or other abstract cognition systems.

Secondly, it is argued that the semantic HTM approach also lends itself naturally to the construction of a motoric hierarchy existing alongside the sensory hierarchy – the key differences being that the semantic-motoric HTM hierarchy refers to configuration space rather than visual space, and is linked to an effector hierarchy rather than a perceptual HTM.

Finally, the notion of a goal HTM is presented, whose nodes contain implications of the form "(perceptual) Context & (motoric) Procedure  $\rightarrow$  Goal", where the Goal is (spatiotemporally bounded) sub-goal of an overall system goal.

A "tripartite network" consisting of interlinked semantic perceptual, motoric and goal hierarchies could be pursued as a holistic AGI approach, or may be connected with more abstract semantic networks and associated cognitive processes to form an integrative AGI approach (such as is done in the OpenCog AGI design).

## 1 Introduction

Hierarchical Temporal Memory is a powerful approach to sensory data processing, exemplified in the HTM (Jeff Hawkins and Dileep George), DeSTIN and HDRN (Itamar Arel), Vicarious Systems (Dileep George) systems and also the joint work of Marek Bundzel and Shuji Hashimoto, and others. HTM presents an attractive combination of intuitive resemblance to the human visual cortex, and practicality in terms of contemporary computer programming and algorithmics.

However, it appears the standard HTM spatiotemporal hierarchy may be too rigid to conveniently and efficiently support the full generality of processing carried out in the human visual system, and needed for human-level vision processing in AGI systems. Furthermore, when one tries to generalize the HTM approach to encompass action as well as perception, one quickly realizes that the HTM hierarchy as conventionally defined is not workable, and a more flexible and abstract approach is needed.

With this in mind, we describe here an extension of the HTM hierarchy that we call "Semantic HTM", which broadens the scope of perceptual processing easily achievable, and also extends more easily to encompass motoric as well as perceptual processing. Finally we present a tripartite HTM architecture, including perceptual, motoric and goal hierarchies, which may be used as an autonomous approach to AGI, or else incorporated with more abstract cognitive AI components to form an integrative system such as OpenCog.

These are initial notes without too much explanation and without any references – reader beware! Familiarity with DeSTIN will be helpful for understanding.

## 2 Semantic HTM for Perception Processing

In the standard HTM hierarchy (here called a "perceptual HTM"), a node  $N$  on level  $k$  (considering level 1 as the bottom) corresponds to a spatiotemporal region  $S$  with size  $s_k$  ( $s_k$  increasing monotonically and usually exponentially with  $k$ ); and, has children on level  $k - 1$  corresponding to spatiotemporal regions that collectively partition  $S$ . For example, a node on level 3 might correspond to a  $16 \times 16$  pixel region  $S$  of 2D space over a time period of 10 seconds, and might have 4 level 2 children corresponding to disjoint  $4 \times 4$  regions of 2D space over 10 seconds, collectively composing  $S$ .

This kind of hierarchy is very effective for recognizing certain types of visual patterns. However it is cumbersome for recognizing some other types of patterns, e.g. the pattern that a face typically contains two eyes beside each other, but at variable distance from each other.

One way to remedy this deficiency is to extend the definition of the hierarchy, so that nodes do not refer to fixed spatial or temporal positions, but only to *relative* positions. In this approach, the internals of a node are basically the same as in an HTM, and the correspondence of the nodes on level  $k$  with regions of size  $s_k$  is retained, but the relationships between the nodes are quite different. For instance, a variable-position node of this sort could contain several possible 2D pictures of an eye, but be nonspecific about where the eye is located in the 2D input image.

Figure ?? depicts this "semantic-perceptual HTM" idea heuristically, showing part of a semantic-perceptual HTM indicating the parts of a face, and also the connections between the semantic-perceptual HTM, a standard perceptual HTM, and a higher-level cognitive semantic network like OpenCog's Atomspace.

More formally, in the suggested "semantic-perceptual HTM" approach, a node  $N$  on level  $k$ , instead of pointing to a set of level  $k - 1$  children, points to a small (but not necessarily connected) *semantic network*, such that

- the nodes of the semantic network are (variable-position) level  $k - 1$  nodes
- the edges of the semantic network possess labels representing spatial or temporal relationships, such as *horizontally\_aligned*, *vertically\_aligned*, *right\_side*, *left\_side*, *above*, *behind*, *immediately\_right*, *immediately\_left*, *immediately\_above*, *immediately\_below*, *after*, *immediately\_after* [these

are illustrative example relationships, not an exhaustive list of possibly useful relationships]

- the edges may also be weighted either with numbers or probability distributions, indicating the quantitative weight of the relationship indicated by the label

So for example, a level 3 node could have a child network of the form `horizontally_aligned( $N_1, N_2$ )` where  $N_1$  and  $N_2$  are variable-position level 2 nodes. This would mean that  $N_1$  and  $N_2$  are along the same horizontal axis in the 2D input but don't need to be immediately next to each other. Or one could say, e.g. `on_axis_perpendicular_to( $N_1, N_2, N_3, N_4$ )`, meaning that  $N_1$  and  $N_2$  are on an axis perpendicular to the axis between  $N_3$  and  $N_4$ . It may be that the latter sort of relationship is fundamentally better in some cases, because *horizontally\_aligned* is still tied to a specific orientation in an absolute space, whereas *on\_axis\_perpendicular\_to* is fully relative. But it may be that both sorts of relationship are useful.

## 2.1 Learning Semantic HTMs

This sort of semantic HTM is much more flexible representationally than the standard HTM, and also, obviously, presents more difficult learning problems. The semantic relationships must be learned, along with the learning occurring within each node. Development of learning algorithms for semantic HTMs is an open problem, but we do have some basic ideas about how it may be done.

First of all, it would seem that, for instance, the DeSTIN learning algorithms could straightforwardly be utilized in the semantic HTM case, once the local semantic networks involved in the network are known. So at least for some HTM designs, the problem of learning the semantic networks may be decoupled somewhat from the learning occurring inside the nodes. DeSTIN nodes deal with clustering of their inputs, and calculation of probabilities based on these clusters (and based on the parent node states). The difference between the semantic HTM and the traditional DeSTIN HTM has to do with what the inputs are.

Regarding learning the semantic networks themselves, one relatively straightforward approach would be to *data mine them from a standard HTM*. That is, if one runs a standard HTM on a stream of inputs, one can then run a

frequent pattern mining algorithm to find semantic networks (using a given vocabulary of semantic relationships) that occur frequently in the HTM as it processes input. A subnetwork that is identified via this sort of mining, can then be grouped together in the semantic HTM, and a parent node can be created and pointed to it.

Also, the standard HTM can be searched for frequent patterns involving the clusters (referring to DeSTIN here, where the nodes contain clusters of input sequences) inside the nodes in the semantic HTM. Thus, in the "semantic DeSTIN" case, we have a feedback interaction wherein:

1. the standard HTM is formed via processing input
2. frequent pattern mining on the standard HTM is used to create subnetworks and corresponding parent nodes in the semantic HTM
3. the newly created nodes in the semantic HTM get their internal clusters updated via standard DeSTIN dynamics
4. the clusters in the semantic nodes are used as seeds for frequent pattern mining on the standard HTM, returning us to Step 2 above

## **2.2 Using the Semantic HTM to Bias the Perceptual HTM**

After the semantic HTM is formed via mining the perceptual HTM, it may be used to bias the further processing of the perceptual HTM. For instance, in DeSTIN each node carries out probabilistic calculations involving knowledge of the prior probability  $P(o)$  of the "observation"  $o$  coming into that node over a given interval of time. In the current DeSTIN version, this prior probability is drawn from a uniform distribution, but it would be more effective to draw the prior probability from the semantic network – observations matching things represented in the semantic network would get a higher prior probability. One could also use subtler strategies such as using imprecise probabilities in DeSTIN, and assigning a greater confidence to probabilities involving observations contained in the semantic network.

## **2.3 Perception-Cognition Interfacing**

Finally, we note that the nodes and networks in the semantic HTM may naturally be linked into the nodes and links in a semantic network such as

OpenCog’s AtomSpace. This allows us to think of the semantic HTM as a kind of bridge between the standard HTM and the cognitive layer of an AI system. In an advanced implementation, the cognitive network may be used to suggest new relationships between nodes in the semantic HTM, based on knowledge gained via inference or language.

### 3 Semantic HTM for Motor and Sensorimotor Processing

The notion that both perception and action are accomplished by hierarchical architectures is far from novel; the core intuition, in the context of visually-guided human actions, is depicted in Figure ???. Here we address this aspect of intelligence via briefly considering a different kind of semantic HTM – one that focuses on movement rather than sensation.

In this case, rather than a 2D or 3D visual space, one is dealing with an  $n$ -dimensional *configuration space* (C-space). This space has one dimension for each degree of freedom of the agent in question. The more joints with more freedom of movement an agent has, the higher the dimensionality of its configuration space.

Using the notion of configuration space, one can construct a *semantic-motoric HTM hierarchy* analogous to the semantic-perceptual HTM hierarchy. However, if one does this in the standard HTM approach, one finds a significant problem – the curse of dimensionality. A square of side 2 can be tiled with 4 squares of side 1, but a 50-dimensional cube of side 2 can be tiled with  $2^{50}$  50-dimensional cubes of side 1. Clearly, if one is to build a hierarchy in configuration space analogous to the HTM hierarchy in perceptual space, some sort of sparse hierarchy is necessary.

There are many ways to build a sparse hierarchy of this nature, but one interesting way is to follow the semantic approach outlined above. This solves the curse of dimensionality, in principle at least. One has a hierarchy where the nodes on level  $k$  represent motions that combine the motions represented by nodes on level  $k-1$ . In this case the most natural semantic label predicates would seem to be things like *simultaneously*, *after*, *immediately\_after*, etc. So a level  $k$  node represents a sort of "motion plan" corresponded by chaining together (serially and/or in parallel) the motions encoded in level  $k-1$  nodes.

Also, it seems there may be even more use for overlapping regions in

the motor case than in the perceptual case. Overlapping regions of C-space correspond to different complex movements that share some of the same component movements, e.g. if one is trying to slap one person while elbowing another, or run while kicking a soccer ball forwards.

It is interesting how the semantic HTM approach reveals perception and motor control to have essentially similar hierarchical structures – a parallel that is much less clear with the traditional HTM approach and its fixed-position nodes.

Just as the semantic-perceptual HTM is naturally aligned with a traditional perceptual HTM, similarly a motoric-perceptual HTM may be naturally aligned with a "motor HTM" of sorts – though a hierarchical temporal motor memory looks quite different from perceptual HTMs like DeSTIN. A flavor of concrete motoric hierarchies in robotics is given in Figure ?? which depicts the portion of a motoric hierarchy corresponding to a robot arm. This sort of hierarchy is intrinsically spatiotemporal because each individual action of each joint of an actuator like an arm is intrinsically bounded in space and time. A more ambitious motoric hierarchy is depicted in Figure ??, which shows how perceptual and motoric hierarchies are constructed and aligned in James Albus's architecture for intelligent automated vehicles.

Figure ?? illustrates the potential alignment between a semantic-motoric HTM and a purely motoric hierarchy. In the figure, the motoric hierarchy is assumed to operate somewhat like DeSTIN, with nodes corresponding to (at the lowest level) individual servomotors, and (on higher levels) natural groupings of servomotors. The node corresponding to a set of servos is assumed to contain centroids of clusters of trajectories through configuration space. The task of choosing an appropriate action is then executed by finding the appropriate centroids for the nodes. Note an asymmetry between perception and action here. In perception the basic flow is bottom-up, with top-down flow used for modulation and for "imaginative" generation of percepts. In action, the basic flow is top-down, with bottom-up flow used for modulation and for imaginative, "fiddling around" style generation of actions. The semantic-motoric hierarchy then contains abstractions of the C-space centroids from the motoric hierarchy – i.e., actions that bind together different C-space trajectories that correspond to the same fundamental action carried out in different contexts or under different constraints. Similarly to in the perceptual case, the semantic hierarchy here serves as a glue between lower-level function and higher-level cognitive semantics.

## 4 Connecting the Perceptual and Motoric Hierarchies with a Goal Hierarchy

Finally, it's interesting to investigate ways of connecting perceptual and motoric hierarchies directly (rather than, say, via means of a separate cognitive network). One way to do this is to envision a "goal HTM" bridging the perceptual and motoric HTMs. The goal HTM would be a "semantic HTM" loosely analogous to the perceptual and motoric semantic HTMs. Each node in the goal HTM would contain implications of the form "Context & Procedure  $\rightarrow$  Goal", where Goal is one of the AI systems overall goals or a subgoal thereof, and Context and Procedure refer to nodes in the perceptual and motoric semantic HTMs respectively.

For instance, a goal HTM node might contain an implication of the form "I perceive my hand is near object X & I grasp object X  $\rightarrow$  I possess object X." This would be useful if "I possess object X" was a subgoal of some higher-level system goal, e.g. if X were a food object and the system had the higher-level goal of obtaining food.

To the extent that the system's goals can be decomposed into hierarchies of progressively more and more spatiotemporally localized subgoals, this sort of hierarchy will make sense, leading to a tripartite hierarchy as loosely depicted in Figure ???. One could attempt to construct an overall AGI approach based on a tripartite hierarchy of this nature, counting on the upper levels of the three hierarchies to come together dynamically to form an integrated cognitive network, yielding abstract phenomena like language, self, reasoning and mathematics. On the other hand, one may view this sort of hierarchy as a portion of a larger integrative AGI architecture, containing a separate cognitive network, with a less rigidly hierarchical structure and less of a tie to the spatiotemporal structure of physical reality. The latter view is the one we are primarily taking within the OpenCog AGI approach, viewing perceptual, motoric and goal hierarchies as "lower level" subsystems connected to a "higher level" system based on the OpenCog AtomSpace and centered on cognitive processes like probabilistic inference, evolutionary program learning, conceptual blending, etc.

Learning of the subgoals and implications in the goal hierarchy is of course a complex matter, which may be addressed via a variety of algorithms, including online clustering (for subgoals or implications) or supervised learning (for implications, the "supervision" being purely internal and provided by goal

or subgoal achievement).

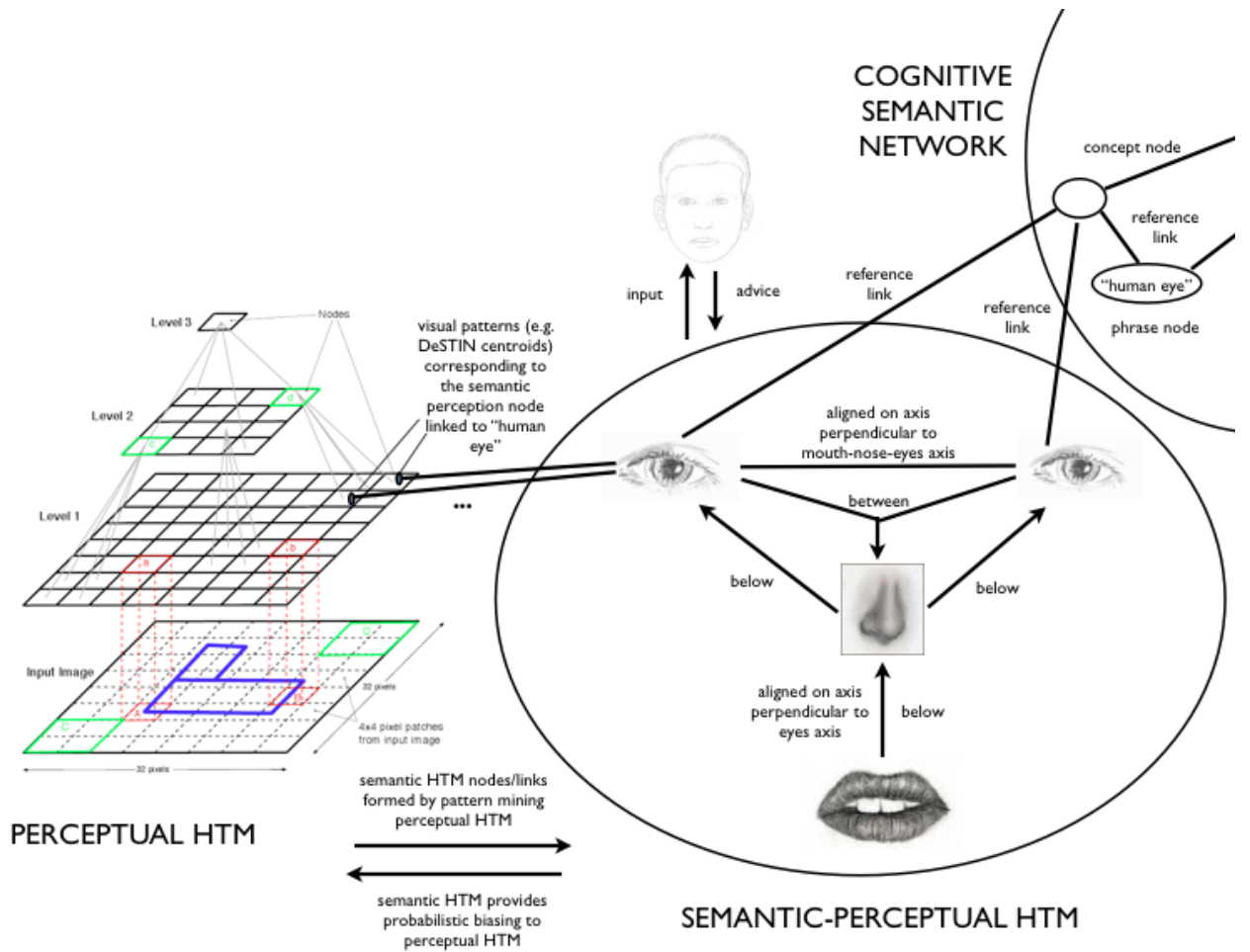


Figure 1: Simplified depiction of the relationship between a semantic-perceptual HTM, a traditional perceptual HTM (like DeSTIN), and a cognitive semantic network (like OpenCog’s AtomSpace). The perceptual HTM shown is unrealistically small for complex vision processing (only 4 layers), and only a fragment of the semantic-perceptual HTM is shown (a node corresponding to the category face, and then a child network containing nodes corresponding to several components of a typical face). In a real semantic-perceptual HTM, there would be many other nodes on the same level as the face node, many other parts to the face subnetwork besides the eyes, nose and mouth depicted here; the eye, nose and mouth nodes would also have child subnetworks; there would be link from each semantic node to centroids within a large number of perceptual nodes; and there would also be many nodes not corresponding clearly to any single English language concept like eye, nose, face, etc.

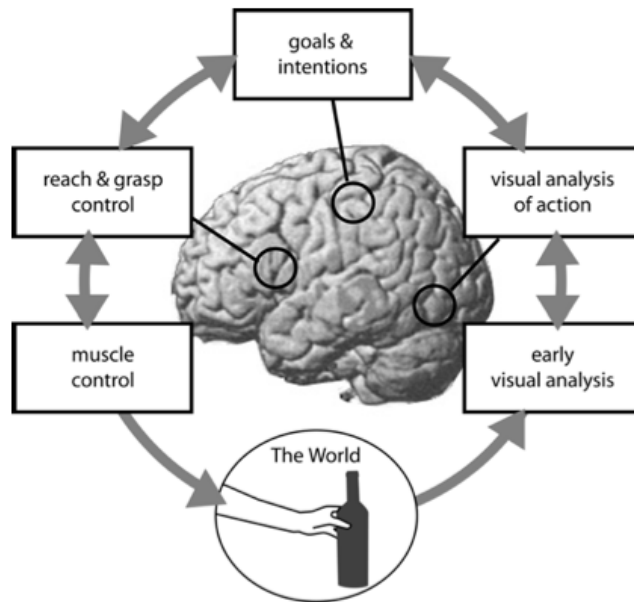


Figure 2: Simplified depiction of the relationship between perceptual and motor hierarchies in human intelligence, taken from *The motor hierarchy: from kinematics to goals and intentions* by Antonia Hamilton and Scott Grafton. In the approach sketched here, the goals and intentions are embodied in a cognitive system like OpenCog, the perceptual hierarchy is a coupling of a perceptual HTM and a semantic-perceptual HTM, and the motor hierarchy is a coupling of a motoric HTM and a semantic-motoric HTM

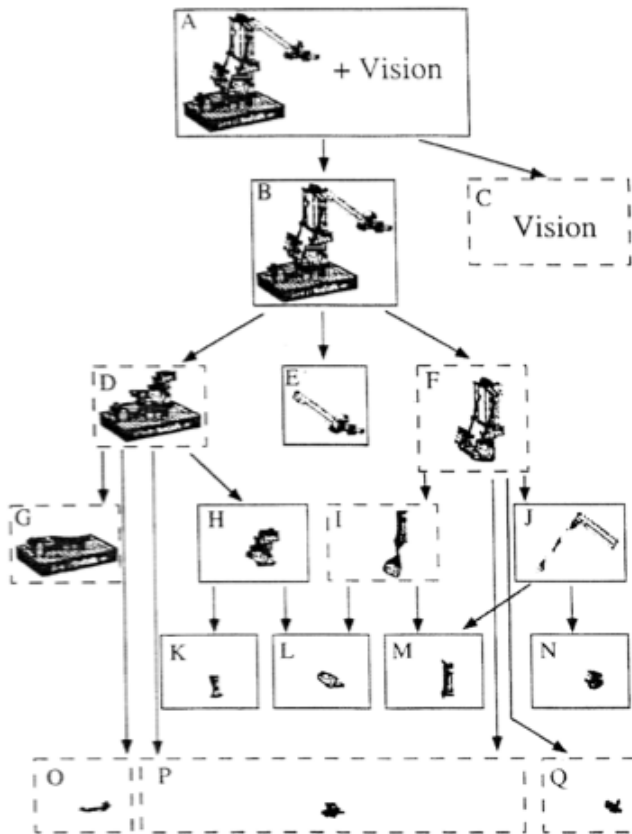


Figure 3: Motor control hierarchy associated with a simple robot arm, taken from *A Hierarchical Intelligent Controller for a Three Link Robot Arm* by Lentz and Acar. In a robot containing an arm plus other actuators, this would be part of the motor control hierarchy, existing in parallel with hierarchies corresponding to other body parts, and supervised by higher-level nodes coordinating the various body parts together.

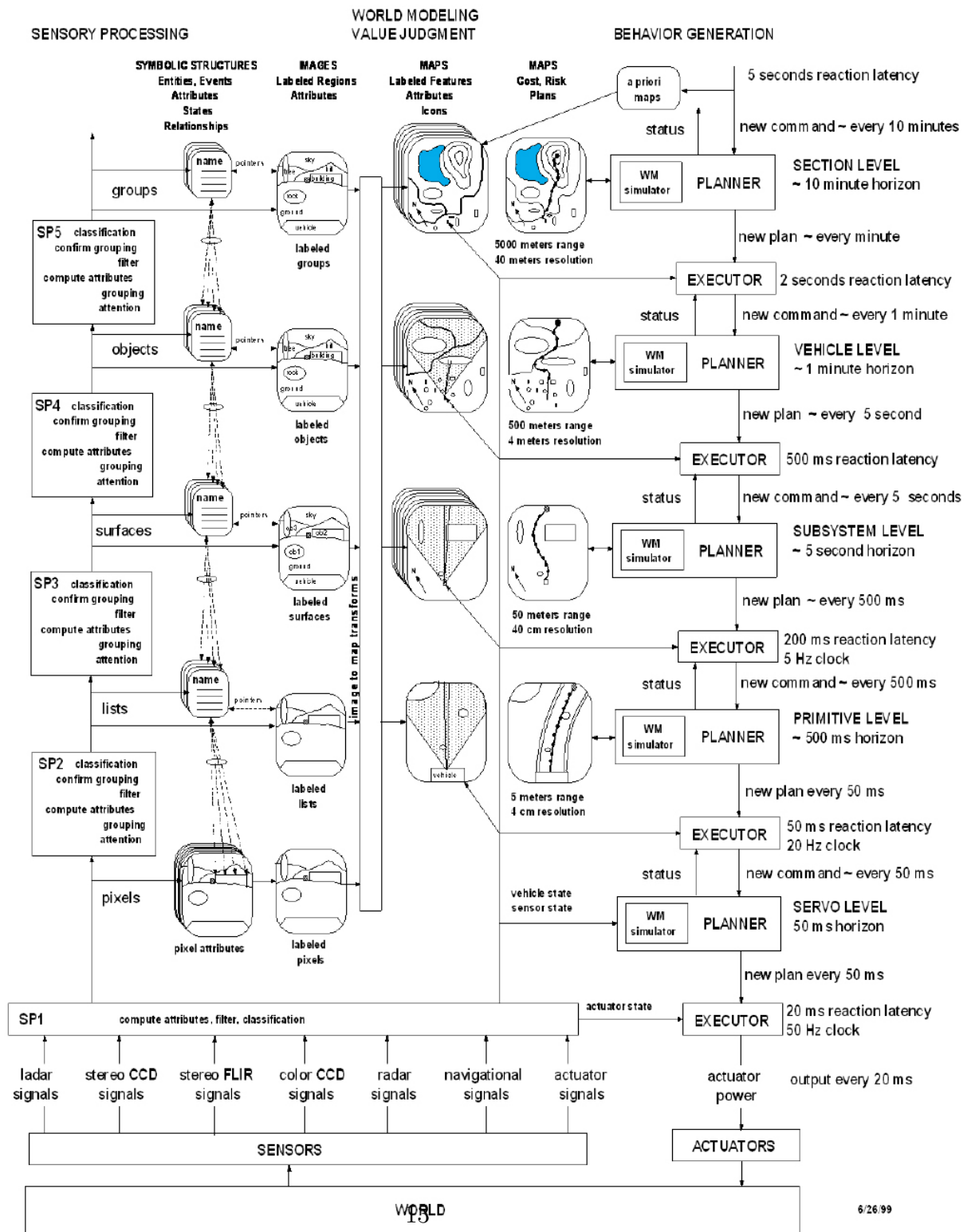


Figure 4: Hierarchical perceptual, motoric and goal architecture used in James Albus's 4D-RCS automated intelligent vehicle control system

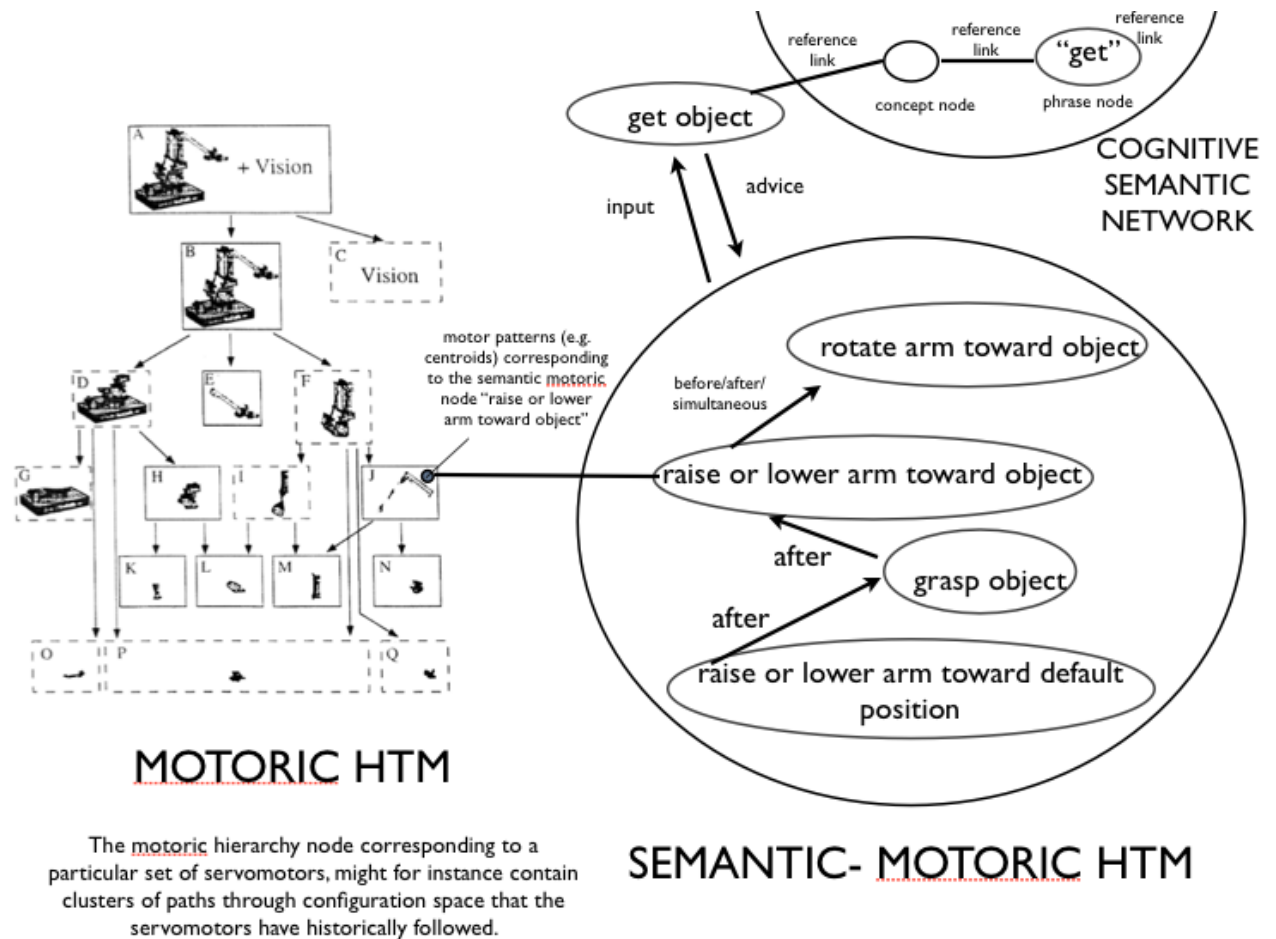


Figure 5: Simplified depiction of the relationship between a semantic-motoric HTM, a motor control hierarchy (illustrated by the hierarchy of servos associated with a robot arm), and a cognitive semantic network (like OpenCog's AtomSpace). Only a fragment of the semantic-motoric HTM is shown (a node corresponding to the "get object" action category, and then a child network containing nodes corresponding to several components of the action). In a real semantic-motoric HTM, there would be many other nodes on the same level as the get-object node, many other parts to the get-object subnetwork besides the ones depicted here; the subnetwork nodes would also have child subnetworks; there would be link from each semantic node to centroids within a large number of motoric nodes; and there might also be many nodes not corresponding clearly to any single English language concept like "grasp object" etc.

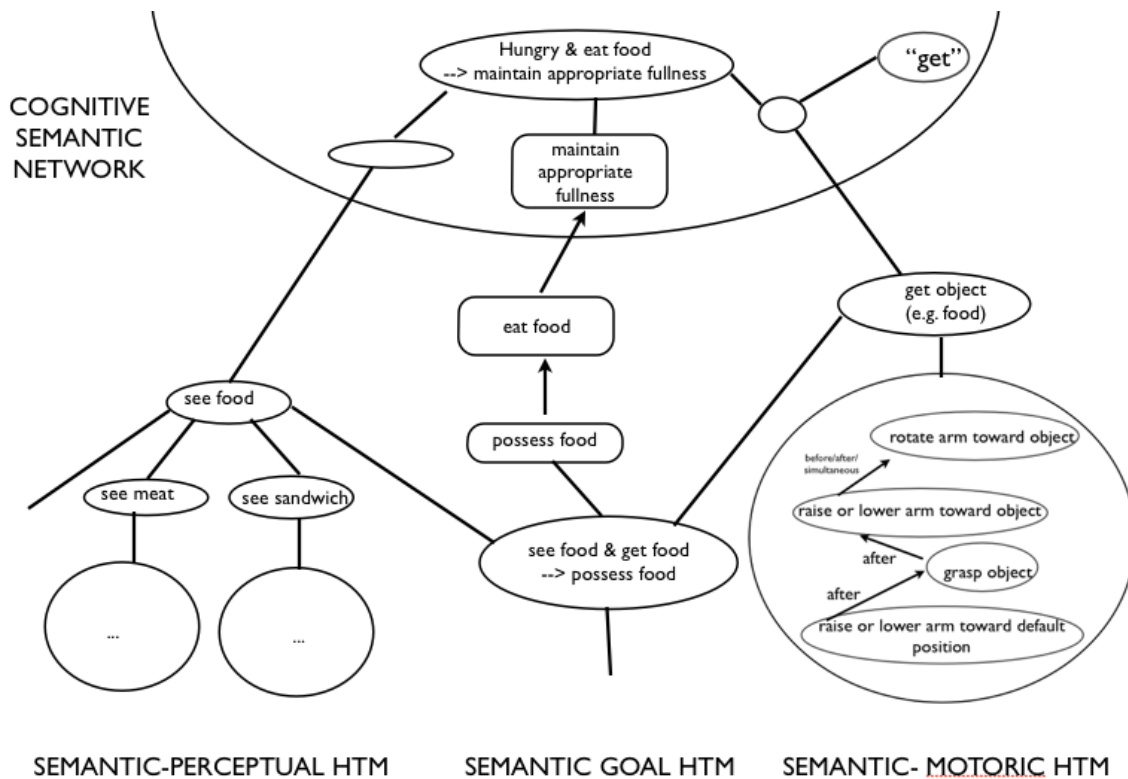


Figure 6: Illustration of the proposed interoperation of perceptual, motoric and goal semantic HTMs. The diagram is simplified in many ways, e.g. only a handful of nodes in each hierarchy is shown (rather than the whole hierarchy), and lines without arrows are used to indicate bidirectional arrows, and nearly all links are omitted. The purpose is just to show the general character of interaction between the components in a simplified context.