Anticipative Control of Voluntary Action: Towards a Computational Model

Pascal Haazebroek[1] and Bernhard Hommel[2,*]

[1] Leiden University
Cognitive Psychology Unit & Leiden Institute for Brain and Cognition
Leiden, The Netherlands
[2] Leiden University
Department of Psychology
Cognitive Psychology Unit
Wassenaarseweg 52
2333 AK Leiden, The Netherlands
hommel@fsw.leidenuniv.nl

Abstract. Human action is goal-directed and must thus be guided by anticipations of wanted action effects. How anticipatory action control is possible and how it can emerge from experience is the topic of the ideomotor approach to human action. The approach holds that movements are automatically integrated with representations of their sensory effects, so that reactivating the representation of a wanted effect by “thinking of it” leads to a reactivation of the associated movement. We present a broader theoretical framework of human perception and action control—the Theory of Event Coding (TEC)—that is based on the ideomotor principle, and discuss our recent attempts to implement TEC by means of a computational model (HiTEC) to provide an effective control architecture for artificial systems and cognitive robots.

Human behavior is commonly proactive rather than reactive. That is, people do not await particular stimulus events to trigger certain responses but, rather, carry out planned actions to reach particular goals. Planning an action ahead and carrying it out in a goal-directed fashion requires prediction and anticipation: in order to select an action that is suited to reach a particular goal presupposes knowledge about relationships between actions and effects, that is, about which goals can be realized by what action. Under some circumstances this knowledge might be generated ad hoc. For instance, should your behavior ever make a flight attendant to drop you by parachute in a desert, your previously acquired knowledge may be insufficient to select among reasonable action alternatives, so you need to make ad hoc predictions to find out where to turn to. But fortunately, most of the situations we encounter are much more familiar and, thus, much easier to deal with. We often have a rough idea about what actions may be suitable under a given goal and in a particular context, simply because we have experience: we have had and reached the same or similar goals and acted in the same or similar situations before.

How experience with one's own actions generates knowledge that guides the efficient selection of actions, and how humans carry out voluntary actions in

* Corresponding author.

G. Pezzulo et al. (Eds.): ABiALS 2008, LNAI 5499, pp. 31–47, 2009.
© Springer-Verlag Berlin Heidelberg 2009

general, was the central issue in ideomotor approaches to human action control. Authors like Lotze (1852), Harless (1861), and James (1890) were interested in the general question of how the mere thought of a particular action goal can eventually lead to the execution of movements that reach that goal in the absence of any conscious access to the responsible motor processes (*executive ignorance*). Key to the theoretical conclusion they came up with was the insight that actions are means to generate perceptions (of wanted outcomes) and that these perceptions can be anticipated. If there would be an associative mechanism that integrates motor processes (m) with representations of the sensory effects they produce (e), and if the emerging association between movements and effect representations would be bidirectional (m←→e), reactivating the representation of the effect by voluntarily "thinking of it" may suffice to reactivate the associated motor processes (e→m). In other words, integrating movements and their sensory consequences provides a knowledge base that allows for selecting actions according to their anticipated outcomes—for anticipative action control that is.

After a flowering period in the second half of the 19th century ideomotor approaches were effectively eliminated from the scientific stage (Prinz, 1987; Stock & Stock, 2004). A major reason for that was the interest of ideomotor theoreticians in conscious experience and the relationship between conscious goal representations and unconscious motor behavior, a topic that did not meet scientific criteria in the eyes of the behaviorist movement gaining power in the beginning of the 20th century (cf., Thorndike, 1913). Starting with an early resurrectional attempt by Greenwald (1970), ideomotor ideas have recently regained scientific credibility and explanatory power however. In their Theory of Event Coding (TEC), Hommel, Müsseler, Aschersleben, and Prinz (2001) have even suggested that the ideomotor principle may represent a firm base on which a comprehensive theory of human perception and anticipatory action control can be built. In the following, we will elaborate on what such a theory may look like. In particular, we will briefly discuss the basic principles and basic assumptions of TEC and then go on to describe our recent attempts to implement these principles and assumptions by means of a computational model of human perception and action control—a model we coined HiTEC (Haazebroek & Hommel, submitted).

1 TEC

The core idea underlying TEC (Hommel et al., 2001) is that perception and action are in some sense the same thing and must therefore be cognitively represented in the same way—the notion of *common coding* (Prinz, 1990). According to the ideomotor principle, action consists in intentionally producing wanted effects, that is, in the execution of motor processes for the sake of creating particular sensory events. In contrast to action, perception is commonly conceived of as the passive registration of sensory input. However, Hommel et al. (2001) argue that this conception is incorrect and misleading, as sensory input is commonly actively produced (Dewey, 1896; Gibson, 1979). For instance, even though visual perception needs light hitting the retina, we actively move our eyes, head, and body to make sure that our retina is hit by the light that is reflecting the most interesting and informative events. That is, we actively search for the information we are interested in and move our receptive surfaces to optimize the intake of that information. This is even more obvious for the tactile sense, as almost nothing would be perceived by touch without systematically moving the sensor surface across the objects of interest. Hence, we perceive by executing motor processes for the sake of creating particular sensory events. Obviously, this is exactly the way we just defined action, which implies that action and perception are one process.

The second central assumption of TEC is that cognitive representations are *composites* of feature codes (Hommel, 2004). Our brain does not represent events through individual codes or neurons but by widely distributed feature networks. For instance, the visual cortex consists of numerous representational maps coding for various visual features, such as color, orientation, shape, or motion (DeYoe & Van Essen, 1988) and similar feature maps have been reported for other modalities. Likewise, action plans are composites of neural networks coding for various action features, such as the direction, force, or distance of manual actions (Hommel & Elsner, 2009). One implication of the assumption that cognitive event representations are composites is that binding operations are necessary to integrate the codes referring to the same event, and another is that different events can be related to, compared with, or confused with each other based on the features they do or do not share. For instance, TEC implies that stimuli and responses can be similar to each other, in the sense that the binding representing the stimulus and the binding representing the response can include the same features, such as location or speed, and can thus prime each other (which for instance explains effects of stimulus-response compatibility) or interact in other ways.

The third main assumption of TEC is that the cognitive representations that underlie perception and action planning code for *distal* but not *proximal* aspects of the represented events (Prinz, 1992). In a nutshell, this means that perceived and produced events are coded in terms of the features of the external event *as external event* (i.e., as objectively or inter-subjectively definable) but not with respect to the specifics of the internal processing, such as retinal or cortical coding characteristics, or particular muscle parameters. This terminology goes back to Heider (1926, 1930), who discussed the problem that our conscious experience refers to objective features of visual objects (the distal attributes), even though the intermediate processing steps of the physical image on the retina and the physiological response to it (the proximal attributes) are not fully determined by the distal attributes. Brunswik (1944) extended this logic to action and pointed out that goal representations refer to distal aspects of the goal event and, thus, do not fully determine the proximal means to achieve it.

To summarize, TEC assumes that perceived events are represented by activating and integrating feature codes—codes that represent the distal features of the event. Given that perceptions are actively produced, these bindings are likely to also include action features, that is, codes that represent the features of the action used to produce that perception. In turn, action plans are integrated bindings of codes representing the distal features of the action. As actions are carried out to create sensory events, action plans also comprise of feature codes referring to these events. In other words, both perceived and produced events are represented by sensorimotor bindings or "event files" (Hommel, 2004). However, not all features of a perceived or a produced event are relevant in a particular context. To account for that, TEC assumes that feature codes are "intentionally weighted" according to the goal or task at hand. For instance, if you are searching for a particular color, or if what matters for your actions is the location of your fingertip, color and location codes would be weighted higher, respectively, and thus affect perception and action planning more strongly. TEC was very helpful in interpreting and integrating available findings in a coherent manner, as well as in stimulating numerous experiments and studies on various topics and perception-action phenomena. However, as Hommel et al. (2001) pointed out, TEC only provides a general framework and the theoretical concepts needed to get a better understanding of higher level perception, action, and their relationship. Deeper insight and theoretical advancement calls for more detail and additional assumptions. To meet this challenge we began developing HiTEC, a computational implementation of TEC's basic principles and assumptions. In the following, we provide a brief overview of the main strategies guiding our implementation, but refer to Haazebroek and Hommel (submitted) for a broader treatment.

2 HITEC

HiTEC (Haazebroek & Hommel, submitted) is an attempt to translate the theoretical framework of TEC (Hommel et al, 2001) into a runnable computational model. Our ambition is to develop a broad, cognitive architecture that can account for a variety of empirical effects related to stimulus-response translation and that can serve as a starting point for a novel control architecture for cognitive robots in the PACO-PLUS project (www.paco-plus.org).

From a modeling perspective TEC provides a number of constraints; some of them enforce structural elements while others impose the existence of certain processes. First, we describe the general structure of HiTEC. Next, we elaborate on the processes operating on this structure, following the two-stage model (Elsner and Hommel, 2001) for the acquisition of voluntary action control. Finally, we discuss how the mechanisms of HiTEC might operate in a real life scenario and show that anticipation plays a crucial role in quickly generating and controlling appropriate responses.

3 HITEC's Structure and Representations

HiTEC is architected as a connectionist network model that uses the basic building blocks of parallel distributed processing (PDP; e.g., McClelland, 1992; Rumelhart, Hinton, & McClelland, 1986). In a PDP network model processing occurs through the interactions of a large number of interconnected elements called units or nodes. Nodes may be organized into higher structures, called modules, each containing a number of nodes. Modules may be part of a larger processing pathway. Pathways may interact in the sense that they can share common modules.

Each node has an activation value indicating local activity. Processing occurs by propagating activity through the network; that is, by propagating activation from one node to the other, via weighted connections. When a connection between two nodes is positively weighted, the connection is excitatory and the nodes will increase each other's activation. When the connection is negatively weighted, it is inhibitory and the nodes will reduce each other's activation. Processing starts when one or more nodes receive some sort of external input. Gradually, node activations will rise and propagate through the network while interactions between nodes control the flow of processing. Some nodes are designated output nodes. When activations of these nodes reach a certain threshold (or when the time allowed for processing has passed), the network is said to produce the corresponding output(s).

In HiTEC, the elementary units are codes. As illustrated in Figure 1, codes are organized into three main systems: the sensory system, the motor system and the common coding system. Each system will now be discussed in more detail.

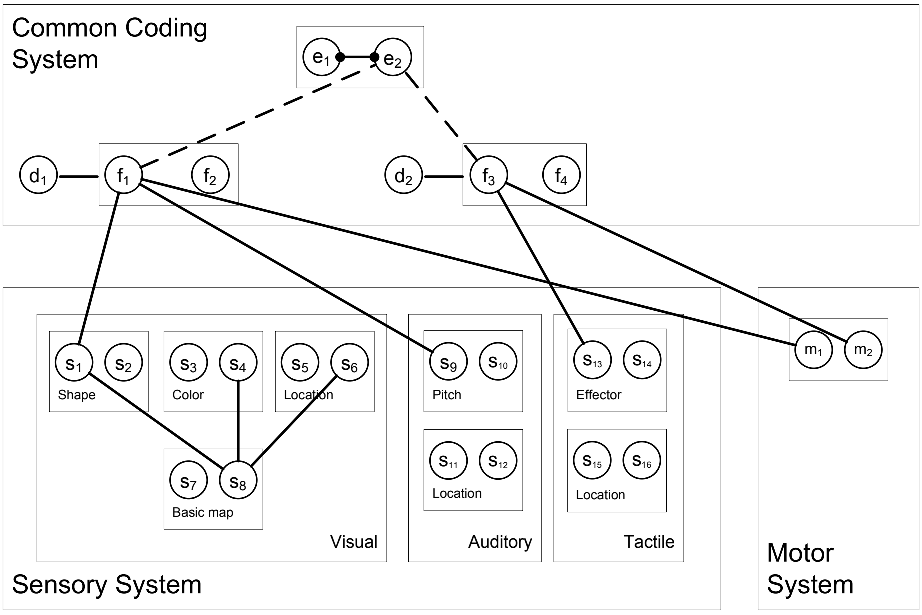

Fig. 1. General architecture of HiTEC

3.1 Sensory System

As already mentioned, the primate brain encodes perceived objects in a distributed fashion: different features are processed and represented across different cortical maps (e.g., Cowey, 1985; DeYoe & Van Essen, 1988). In HiTEC, different modalities (e.g., visual, auditory) and different dimensions within each modality (e.g., visual color and shape, auditory location and pitch) are processed and represented in different sensory maps. Each sensory map is a module containing a number of sensory codes that are responsive to specific sensory features (e.g., a specific color or a specific pitch). Note that Figure 1, shows only two sensory codes per map for clarity.

In the visual brain, there are two major parallel pathways (Milner & Goodale, 1995) that follow a common preliminary basic feature analysis step. The ventral pathway is seen as crucial for object recognition and consists of a hierarchy of sensory maps coding for increasingly complex features (from short line segments in the lower maps to complex shapes in higher maps) and increasingly large receptive field (from a small part of the retina in the lower maps to anywhere on the retina in higher maps). The second pathway, the dorsal pathway, is seen as crucial for action guidance as it loses color and shape information but retains information about contrast, location of objects, and other action-related features.

In HiTEC, a common visual sensory map codes for basic visual parts of perceptual events. This common basic map projects to both the ventral and the dorsal pathways. The ventral pathway consists of sensory maps coding for combinations (such as more specific shapes) or abstractions (e.g., object color). The dorsal pathway is currently simply a sensory map coding for visual location—to be extended for processing other action-related features in a later version of HiTEC.

Distributed processing allows a system to dramatically increase its representational capacity as it no longer requires each combination of features to have its own dedicated representational structure but can rather encode a specific combination on demand in terms of activating a collection of constituting feature structures. On the downside, in typical scenarios, this inevitably results in binding problems (Treisman, 1996). For instance, when multiple objects are perceived and they are both represented in terms of activating the structures coding for their constituting features, how to tell which feature belongs to which object? This clearly calls for an integration mechanism that can tell them apart.

Recent studies in the visual modality have shown that this problem can, partly, be solved by employing local interactions between feed-forward and feed-back processes in the ventral and dorsal pathways (Van der Velde & De Kamps, 2001). It is true that higher ventral sensory maps do not contain information on location and that higher dorsal sensory maps do not contain information on object shape or color, but these pathways can interact using the common basic visual feature map as a visual blackboard (Van der Velde, De Kamps, & Van der Voort van der Kleij, 2004). For instance: when a specific color is activated in a higher sensory map, it can feed back activation to lower sensory maps, thereby modulating the activity of these sensory codes in a way that those codes that code for simple parts of this color are enhanced. This can modulate the processing in the dorsal pathway as well resulting in enhanced activation of those codes in the location map that code for the location(s) of objects of the specified color.

This principle also works the other way round: activating a specific location code in the location map can modulate the sensory codes in the lower sensory maps that code for simple parts at this location. This can modulate the processing in the ventral pathway, resulting in enhanced activation of the more complex or abstract features of the object at the specified location. In HiTEC, this is the way the visual sensory system can be made to enhance the processing of objects with specific features or on a specific location. For now, we assume the following sensory maps in the HiTEC architecture: visual basic features map, visual color map, visual shape map, visual location map, auditory pitch map, auditory location map, tactile effector (i.e., hands or feet) map and tactile location map.

3.2 Motor System

The motor system contains motor codes, referring to proximal aspects of movements. Motor codes can also be organized in maps, following empirical evidence that suggests distributed representations at different cortical locations in the motor domain (e.g., Andersen, 1988; Colby 1998). For example, cortical maps can be related to effector (e.g., eye, hand, arm, foot) or movement type (e.g., grasping, pointing). It makes sense to assume that there is some sort of hierarchical structure as well in motor coding. However, in the present version of HiTEC, we consider only one basic motor map with a set of motor codes. As our modeling efforts in HiTEC evolve, its motor system may be extended further.

It is clear that motor codes, even when structured in multiple maps, can only specify a rough outline of the motor action to be performed as some parameters depend strongly on the environment. For instance, when grasping an object, the actual object location is not represented by a motor code (this would lead to an explosion of the number of necessary motor codes, even for a very limited set of actions). So it makes sense to interpret a motor program as a blueprint of a motor action that needs to be filled in with this specific, on line, information, much like the schemas put forward by Schmidt (1975) and Glover (2004). In our discussion of HiTEC processes we will discuss this issue in more detail.

3.3 Common Coding System

According to TEC both perceived events and action generated events are coded in one common representational domain (Hommel et al, 2001). In HiTEC, this domain is the common coding system that contains common feature codes. Feature codes refer to distal features of objects, people and events in the environment. Example features are distance, size and location, but on a distal, descriptive level, as opposed to the proximal features as coded by the sensory codes and motor codes.

Feature codes may be associated to both sensory codes and motor codes and are therefore truly sensorimotor. They can combine information from different modalities and are in principle unlimited in number. Feature codes are not given but they evolve and change. In HiTEC simulations, however, we usually assume a set of feature codes to be present initially, to bootstrap the process of extracting sensorimotor regularities in interactions with the environment.

Feature codes are contained in feature dimensions. As feature dimensions may be enhanced as a whole, for each dimension an additional dimension code is added that is associated with each feature code within this dimension. Activating this code will spread activation towards all feature codes within this dimension, making them more sensitive to stimulation originating from sensory codes.

3.4 Associations

In HiTEC, codes can become associated, both for short term and for long term. Short term associations between feature codes reflect that these codes 'belong together in the current task or context' and their binding is actively maintained in working memory. In Figure 1, these temporary bindings are depicted as dashed lines. Long term associations can be interpreted as learned connections reflecting prior experience. For now, we do not differentiate between episodic and semantic memory—even though later versions are planned to distinguish between a “literal” episodic memory that stores event files (see below) and a semantic memory that stores rules abstracted from episodic memory (O'Reilly & Norman, 2002). At present, both types of experience are modeled as long term associations between (any kind of) codes and are depicted as solid lines in Figure 1.

3.5 Event file

Another central concept in the theory of event coding is the event file (Hommel, 2004). In HiTEC, the event file is modeled as a structure that temporarily associates to feature codes that 'belong together in the current context' in working memory. The event file serves both the perception of a stimulus as well as the planning of an action. Event files can compete with other event files.

4 HITEC's Processes

How do associations between codes come to be? What mechanisms result of their interactions? And how do these mechanisms give rise to anticipation based, voluntary action control? Elsner and Hommel (2001) proposed a two-stage model for the acquisition of voluntary action control. At the first stage, the cognitive system observes and learns regularities in motor actions and their effects. At the second stage, the system uses the acquired knowledge of these regularities to select and control its actions. For both stages, we now discuss in detail how processes take place in the HiTEC architecture. Next, we discuss some additional process related aspects of the architecture.

Stage 1: Acquiring Action-Effect Associations

The framework of event coding assumes that feature codes are grounded representations as they are derived by abstracting regularities in activations of sensory codes. However, the associations between feature codes and motor codes actually signify a slightly different relation: feature codes encode the (distal) perceptual effect of the action that is executed by activating the motor codes. Following the ideomotor principle, the cognitive system has no innate knowledge of the actual motor action following the activation of a certain motor code. Rather, motor codes need to become associated with their perceptual action effects so that by anticipating these effects, activation can propagate via these associations to those motor codes that actually execute the corresponding movement.

Infants typically start off with a behavioral repertoire based on stimulus-response (SR) reflexes (Piaget, 1952). As the infant exhibits these stimulus-response reflexes, as well as random behaviors (e.g., motor babbling), its cognitive system learns the accompanying response-perceptual effect (RE) regularities that will serve as some sort of database of 'what action achieves what environmental effect'. Following Hommel (1996), we assume that any perceivable action effect is automatically coded and integrated into an action concept, which is, in the HiTEC architecture, an event file consisting of feature codes. Although all effects of an action become integrated automatically, intentional processes do affect the relative weighting of integrated action effects—TEC's intentional-weighting principle.

Taken together, action – effect acquisition is modeled in HiTEC as follows: motor codes m_i are activated, either because of some already existing associations or simply because of network noise. This leads to a change in the environment (e.g., the left hand suddenly touches a cup) which is picked up by sensory codes s_i. Activation propagates from sensory codes towards feature codes f_i. And eventually, these feature codes are integrated into an event file e_i which acts as an action concept. Subsequently, the cognitive system learns associations between the feature codes f_i belonging to this action concept and the motor code m_i that just led to the executed motor action. Crucially, task context can influence the learning of action effects. Not by selecting which effects are associated but by weighting the different effect features. Nonetheless, this is an interactive process that does not exclude unintended but utterly salient action effects to become involved in strong associations as well.

Stage 2: Using Action Effect Associations

Once associations between motor codes and feature codes exist, they can be used to select and plan voluntary actions. Thus, by anticipating desired action effects, feature codes become active. Now, by integrating the feature codes into an action concept, the system can treat the features as constituting a desired state and propagate their activation towards associated motor codes. Crucially, anticipating certain features needs integration to tell them apart from the features that code for the currently observed environment. Once integrated, the system has 'a lock' on these features and can use these features to select the right motor action.

Initially, multiple motor codes m_i may become active as they typically fan out associations to multiple feature codes f_i. However, some motor codes will have more associated features that are also part of the active action concept and some of the m_i - f_i associations may be stronger than others. Taken together, the network will – in PDP fashion – converge towards one strongly activated motor code m_i which will lead to the selection of that motor action.

In addition to the mere selection of a motor action, feature codes also form the actual action plan that specifies (in distal terms) how the action should be executed: namely, in such a way the intended action effects are realized. By using anticipated action effects to choose an action, the action actually is selected because the cognitive system intended this, not because of a reflex to some external stimulus. Thus, in HiTEC, using anticipation is the key to voluntary action.

4.1 Task Context

Task context can modulate both action-effect learning and the usage of these links. This can help focus processing to action alternatives that 'make sense' in the current context. In real life this is necessary as the action alternatives are often rather unconstrained. Task context comes in different forms. One is the overall environment, the scene context in which the interaction takes place. The cognitive system may just have seen other objects in the room, or the room itself, and feature codes that code for aspects of this context may still have some activation. This can, in principle, influence action selection. As episodic and semantic memory links exist as well, this influence may also be less salient: the presence of a certain object might recall memories of previous encounters or similar contexts that influence action selection in the current task.

A task can also be very specific, as given by a tutor or instructor in terms of a verbal description. In HiTEC, it is assumed that feature codes can be activated by means of verbal labels. Thus, when a verbal task is given, this could directly activate feature codes. The cognitive system integrates these codes into an event file that is actively maintained in working memory. For example, when approached with several options to respond differently to, different event files e_i are created for the different options. Due to the mutual inhibitory links between event files, they will compete with each other. Because of the efficiency the cognitive system can now display, one could state that a cognitive reflex has been prepared (Hommel, 2000) that anticipates certain stimuli features. The moment these features are actually perceived, the reflex 'fires' and - by propagating activation to event codes and subsequently to other feature codes - quickly anticipates the correct action effects, which results in the selection and execution of the correct motor action.

4.2 Online vs. Offline Processing

In HiTEC, action selection and action planning are interwoven, but on a distal feature level. This leaves out the necessity of coding every minute detail of the action, but restricts action planning to a ballpark idea of the movement. Still, a lot has to be filled in by on line information. Currently, this falls outside the scope of HiTEC, but one could imagine that by activating distal features, the proximal sensory codes can be top down moderated to 'focus their attention' towards specific aspects of the environment (e.g., visual object location), see Hommel (in press). In addition, actions need still not to be completely specified in advance, as they are monitored and adjusted while they are performed—which in humans seems to be the major purpose of dorsal pathways (Milner & Goodale, 1995)

4.3 Action Monitoring

The anticipated action effects are a trigger for action selection, but also form an expectation of the perceptual outcome of the action. Differences between this expectation and reality lead to adjusting the action on a lower sensorimotor level than is currently modeled in HiTEC. What matters now, is that the feature codes are interacting with the sensory codes, making sure that the generated perception is within the set parameters, as determined by the expected action outcome. If this is not (well enough) the case, the action should be adjusted.

However, when a discrepancy of this expectation drastically exceeds 'adjustment thresholds', it may actually trigger action effect learning (stage 1). Apparently, the action-effect associations were unable to deliver an apt expectation of the actual outcome. Thus, anticipating the desired outcome falsely led to the execution of this action. This may trigger the system to modify these associations, so that the motor codes become associated with the correct action effect features.

Crucially, having anticipations serve as expectations, the system is not forced into two distinct operating modes (learning vs. testing). With anticipation as retrieval cue for action selection and as expectation of the action outcome, the system has the means to self-regulate its learning by making use of the discrepancy between actual effects and these anticipations.

5 Model Implementation

The HiTEC model is implemented using neural network simulation software that facilitates the specification and simulation of interactive networks. In interactive networks, connections are bidirectional and the processing of any single input occurs dynamically during a number of cycles. Each cycle, the network is gradually updated by changing the activation of each node as a result of its interactions with other nodes.

5.1 Code Dynamics

HiTEC aims at a biologically realistic implementation of network dynamics. In the human brain, local interactions between neurons are largely random, but when looking at groups of neurons (i.e., neuron populations) their average activation can be described using mean field approximation equations (Wilson and Cowan, 1972). In HiTEC, a single code is considered to be represented by a neuron population. Its dynamics can therefore be described using differential equations such as:

$$\frac{dA}{dt} = -A + \sum_k w_k F(A_k) + N$$

This equation states that the change in activation A of a code is a result of a decay term and the weighted sum of the outputs of those nodes k that it connects to. Also, each node receives additional random noise input N. Node output is computed using an activation function $F(A)$ that translates node activation into its output as governed by the following logistic function:

$$F(A) = \frac{1}{1 + \exp(-A)}$$

The simulator uses numerical integration to determine the change of activation for each node in each cycle.

5.2 Codes

Currently, in our simulations we hard code all sensory codes including their receptive field specification (e.g., whether a code is responsive to a red or a blue color). Also, feature codes are assumed to exist, as well as all connections between sensory codes and feature codes reflecting prior experience with sensory regularities. In the future it may become an interesting endeavor to learn the grounding of feature codes in terms of proximal sensory codes, possibly by means of self organizing map methods that can be moderated by HiTEC processes (e.g., failing to predict an action outcome may signal relevant novelty and moderate the creation or update of a feature code). Also, for now, we assume a limited set of motor programs that are simply represented by fixed motor codes. Thus, in simulations we currently focus on the interactions between perception and action and how task context influence these interactions, rather than on the grounding of codes per se.

5.3 Action-Effect Learning

Learning action effects is reflected by creating long term connections between feature codes and motor codes. This is currently done by simple associative, Hebbian learning, as described by the following equation:

$$\frac{dw_{ij}}{dt} = \gamma A_j (A_i - w_{ij})$$

Thus, the change of the connection strength is determined by the activation of the nodes *i* and *j* that are connected. This way, feature codes that were activated more strongly will become more strongly connected to the motor code that caused the perceptual effect. Surely, this type of learning is known to be limited but serves our current purposes.

5.4 Short Term Associations and Event File Competition

Crucial in HiTEC is the short term memory component. A task instruction is represented using short term connections between feature codes and event files. In the current set up, an event file is simply a node that is created on demand, as a result of the task instruction, and temporarily connects to those feature codes that were activated by the task instruction (i.e., via verbal labels). An event file has an enhanced baseline activation, reflecting its task relevance. Moreover, event files compete with each other by means of lateral inhibition (i.e., they are interconnected with negative connections) resulting in a winner-take-all mechanism: as activation gradually propagates from feature codes to event files (and back), their activation changes as well. Due to the lateral inhibition, only one event file will stand as the 'winner', while weakening the other event files. This results in selective activation at the feature code level and subsequently in action selection at the motor code level.

5.5 Related Work

We must note that we do advertise the associative learning method used in HiTEC as a competitive alternative to highly specialized machine learning techniques that are traditionally used in classification tasks (e.g., Hiddden Markov Models, Support Vector Machines et cetera) or reward based learning tasks (e.g., Reinforcement learning, Q-learning et cetera). However, we do focus on the context of learning: the interplay between (the coding of) task context and action effect anticipation and perception triggers and mediates learning. In particular, we stress that the cognitive system employs anticipation as reflection of both its learned knowledge so far and its interpretation of the current context. Anticipation can subsequently mediate learning by influencing which features engage in learning (and even further: what features to look for in the sensory input) and how strongly these features may be associated to motor codes, thereby constraining whatever (machine) learning technique used to actually create or change the associations.

Moreover, failing to correctly anticipate an action effect may be a major trigger to update the learned knowledge. In the future we may add this as a reinforcement learning component that drives on biologically plausible reward mechanisms (e.g., dopamine moderated learning).

Finally, we stress that although simulations may be set up in terms of instruction, train and test phases, the HiTEC model itself does not artificially 'switch' between two modes of operation: learning occurs on line as a result of perceiving action effects.

6 Examplary Scenario: Responding to Traffic Lights

In order to clarify the co-operation of the different processes and mechanisms in HiTEC on a functional level, the following example real life scenario is presented: learning to respond to traffic lights. In this example, s_i denotes sensory codes, f_i denotes feature codes and m_i denotes motor codes in the HiTEC architecture. Figure 2 shows a scenario-specific version of the HiTEC architecture.

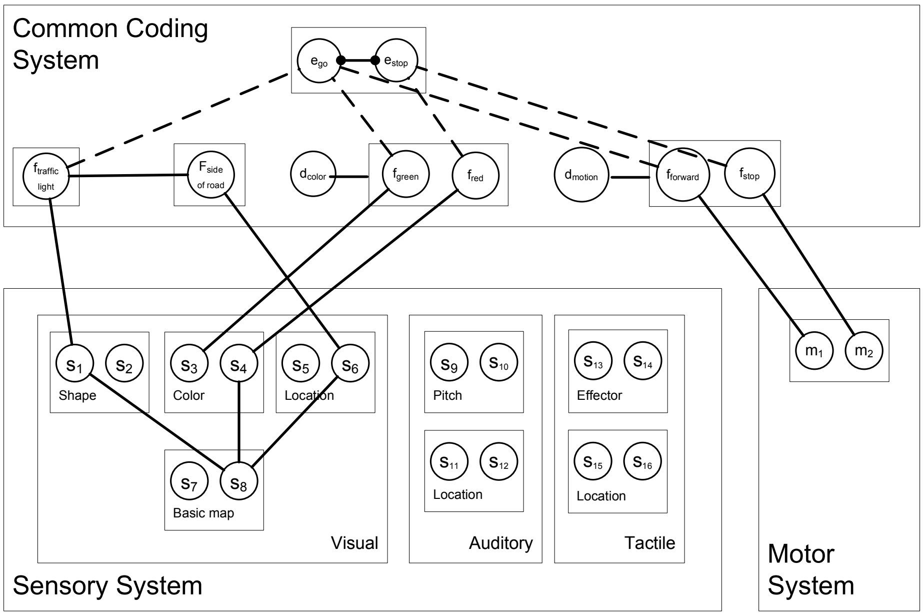

Fig. 2. Learning to respond to traffic lights in HiTEC

6.1 Action Effect Acquisition

Let's say you are a student driver who has never paid attention to the front seat before and this is your first driving lesson. You climb behind the steering wheel and place your feet above the pedals. Now, the instructor starts the car for you and you get the chance of playing around with the pedals. After a while, you get the hang of it: it seems that pressing the right pedal results in a forward movement of the car, and pressing the left one puts the car on hold.

From a HiTEC perspective, you just have tried some motor codes and learned that m_1 (pressing the gas pedal) results in a forward motion, coded by $f_{forward}$ and m_2 in standing still, coded by f_{stop}. In other words: you acquired these particular action-effect associations. Note that we assume that you have been able to walk before, so it is fair to say that $f_{forward}$ and f_{stop} are already present as feature codes in your common coding system.

6.2 Using Action Effect Associations

Now, in your next lesson you actually need to take cross roads. The instructor tells you to pay attention to these colored lights next to the road. When the red light is on, you should stop, and when the green light is on, you can go forward.

In HiTEC, this verbal instruction is modeled as creating two event files that hold short term associations in working memory: $e_{stop\ for\ red\ light}$ for the 'stop' condition, and $e_{go\ at\ green\ light}$ for the 'forward' condition. The event file $e_{stop\ for\ red\ light}$ contains bindings of feature codes f_{red}, $f_{traffic\ light}$, f_{stop} and the event file $e_{go\ at\ green\ light}$ relates to the feature codes f_{green}, $f_{traffic\ light}$, $f_{forward}$.

These event files are activated and their activation spreads to their associated feature codes which will become increasingly receptive for interaction with related sensory codes. In addition to the specific features, the feature dimensions these features are contained in (d_{color}, d_{motion}) are weighted as well. The anticipation of traffic lights also serves as a retrieval cue for prior experience with looking at traffic lights. As traffic lights typically stand at the side of the road, one could expect associations between $f_{traffic\ light}$ and $f_{side\ of\ road}$ to exist in episodic or semantic memory. Consequently, anticipating a traffic light activates $f_{traffic\ light}$ and propagates activation automatically towards $f_{side\ of\ road}$, which makes the system more sensitive to objects located on the side of the road.

Ok, there it goes... you start to drive around, take some turns, and there it is… your very first cross road with traffic lights!

Now, from a HiTEC perspective, the following takes place: the visual scene consists of a plethora of objects, like road signs, other cars, houses and scenery, and of a cross road with traffic lights at the side. The sensory system encodes the registration of these objects by activating the codes in the sensory maps. This leads to the classical binding problem: multiple shapes are registered, multiple colors and multiple locations. However, we now have a top down 'special interest' for traffic lights. As mentioned above, this has resulted in increased sensitivity of the $f_{traffic\ light}$ feature code, that now receives some external stimulation from related sensory codes. Also, from prior experience we look more closely at $f_{side\ of\ road}$ locations in the sensory location maps.

The interaction between this top down sensitivity and the bottom up external stimulation results in an interactive process where the sensory system uses feedback signals to the lower level visual maps where local interactions result in higher activation of those sensory codes that code for properties of the traffic light, including its color. In the visual map for object color, the traffic light color will be more enhanced than colors relating other objects. On the feature code level, the color dimension already was enhanced because of the anticipation of features in the d_{color} dimension, resulting in fast detection of f_{red} or f_{green}.

Meanwhile, the event files $e_{stop\ for\ red\ light}$ and $e_{stop\ for\ red\ light}$ are still in competition. When the sensory system collects the evidence, activation propagates towards feature codes and event codes, quickly converging into a state that where either $f_{forward}$ or f_{stop}

is activated more strongly than the other. This activation is propagated towards the motor codes m_1 or m_2 via associations learned in your first drivers lesson. This results in the selection and execution of the correct motor action.

It is clear that by preparing the cognitive system for perceiving a traffic light color and producing a stop-or-go action allows the system to effectively attend its resources to the crucial sensory input and already pre-anticipate the possible action outcome. This way, upon perceiving the actual traffic light color, the system can quickly respond with the correct motor action.

Luckily, for your safety and that of all your fellow drivers on the road, practicing this task long enough will also result in long term memory bindings between f_{red}, $f_{traffic\ light}$ and f_{stop} that will also be retrieved during action selection and bias you towards pressing the brake pedal, even when no instructor is sitting next to you.

7 Conclusions

We have introduced HiTEC's three main modules: the sensory system, the motor system, and the emergent common coding system. These systems interact with each other. In the common coding system anticipations are formed that have a variety of uses in the architecture, allowing the system to be more flexible and adaptive. In action selection, anticipation acts as a rich retrieval cue for associated motor programs. At the same time, forming this anticipation reflects the specification of an action plan that can be used during action execution.

One of the drawbacks of creating anticipations is that it might not be worth the costs (Butz & Pezzulo, 2008). However, from a real life scenario perspective, the number of possible action alternatives is enormous. Creating anticipations at a distal level seems as a necessity to constrain the system in its actions to select from. Doing this, as we propose in HiTEC, not only aids action selection but also delivers the rudimentary action plan at the same time.

Another concern often mentioned is the inaccuracy of predictions. Following the framework of event coding, events – including action plans – are coded in distal terms that abstract away from the proximal sensory values. Only inaccuracies on the distal level could disturb the use of anticipations in action selection and planning. The feature codes on this distal level are based on sensorimotor regularities that are stable over time. Thus minor inaccuracies in sensors should be relatively easily overcome.

Actions are usually selected and planned in a task context. When forced with different behavioral alternatives to choose from, multiple anticipations of features are created and compete with each other. When features are actually perceived, anticipatory activation quickly propagates to the correct action effects, which results in the selection and execution of the correct motor action.

In action monitoring, anticipation serves as the representation of expected and desired action effects that helps adjusting the movement during action execution. In action evaluation, this expectation acts as a set of criteria for success of the action. If the actual action effect can no longer – on a lower sensorimotor level - be adjusted to fulfill the expected action effect, the existing action-effect associations are considered insufficient and learning is triggered. During action-effect learning, anticipation also may weight the different action effect features in the automatic integration into action concepts, influencing the action-effect association weights.

In conclusion, anticipation plays a crucial role in virtually all aspects of action control within the HiTEC architecture. Just as it does in real life.

Acknowledgments

Support for this research by the European Commission (PACO-PLUS, IST-FP6-IP-027657) is gratefully acknowledged.

References

1. Andersen, R.A.: The neurobiological basis of spatial cognition: Role of the parietal lobe. In: Stiles-Davis, J., Krtichivsky, M., Belugi, U. (eds.) Spatial cognition: Brain bases and development. Erlbaum, Mahwah (1988)
2. Brunswik, E.: Distal focussing of perception: Size constancy in a representative sample of situations. Psychological Monographs 56(1) (1944)
3. Butz, M.V., Pezzulo, G.: Benefits of Anticipations in Cognitive Agents. In: Pezzulo, G., Butz, M.V., Castelfranchi, C., Falcone, R. (eds.) The Challenge of Anticipation. LNCS, vol. 5225, pp. 45–62. Springer, Heidelberg (2008)
4. Colby, C.L.: Action-oriented spatial reference frames in the cortex. Neuron 20, 15–20 (1998)
5. Cowey, A.: Aspects of cortical organization related to selective attention and selective impairments of visual perception: A tutorial review. In: Poster, M.I., Marin, O.S.M. (eds.) Attention and performance XI, pp. 41–62. Erlbaum, Hillsdale (1985)
6. Dewey, J.: The reflex arc concept in psychology. Psychological Review 3, 357–370 (1896)
7. DeYoe, E.A., Van Essen, D.C.: Concurrent processing streams in monkey visual cortex. Trends in Neuroscience 11, 219–226 (1988)
8. Elsner, B., Hommel, B.: Effect anticipation and action control. Journal of Experimental Psychology: Human Perception and Performance 27 (2001)
9. Gibson, J.J.: The ecological approach to visual perception. Houghton Mifflin, Boston (1979)
10. Glover, S.: Separate visual representations in the planning and control of action. Behavioral and Brain Sciences 27, 3–24 (2004)
11. Greenwald, A.: Sensory feedback mechanisms in performance control: With special reference to the ideomotor mechanism. Psychological Review 77, 73–99 (1970)
12. Harless, E.: Der Apparat des Willens. Zeitschrift fuer Philosophie und philosophische Kritik 38, 50–73 (1861)
13. Haazebroek, P., Hommel, B.: HiTEC: A computational model of the interaction between perception and action (submitted)
14. Heider, F.: Thing and medium. Psychological Issues, Monograph 3 (original work published 1959) (1926/1959)
15. Heider, F.: The function of the perceptual system. Psychological Issues, 1959, Monograph, 371–394 (1930/1959) (original work published 1930)
16. Hommel, B.: The cognitive representation of action: Automatic integration of perceived action effects. Psychological Research 59, 176–186 (1996)
17. Hommel, B.: The prepared reflex: Automaticity and control in stimulus-response translation. In: Monsell, S., Driver, J. (eds.) Control of cognitive processes: Attention and performance XVIII, pp. 247–273. MIT Press, Cambridge (2000)

18. Hommel, B.: Event files: Feature binding in and across perception and action. Trends in Cognitive Sciences 8, 494–500 (2004)
19. Hommel, B.: Grounding attention in action control: The intentional control of selection. In: Bruya, B.J. (ed.) Effortless attention: A new perspective in the cognitive science of attention and action. MIT Press, Cambridge
20. Hommel, B., Elsner, B.: Acquisition, representation, and control of action. In: Morsella, E., Bargh, J.A., Gollwitzer, P.M. (eds.) Oxford handbook of human action, pp. 371–398. Oxford University Press, New York (2009)
21. Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W.: The Theory of Event Coding (TEC): A framework for perception and action planning. Behavioral and Brain Sciences 24, 849–937 (2001)
22. James, W.: The principles of psychology. Dover Publications, New York (1890)
23. Lotze, R.H.: Medicinische Psychologie oder die Physiologie der Seele. Weidmann'sche Buchhandlung, Leipzig (1852)
24. McClelland, J.L.: Toward a theory of information processing in graded, random, and interactive networks. In: Meyer, D.E., Kornblum, S. (eds.) Attention and performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience – A Silver Jubilee Volume. MIT Press, Cambridge (1992)
25. Milner, A.D., Goodale, M.A.: The visual brain in action. Oxford University Press, Oxford (1995)
26. O'Reilly, R.C., Norman, K.A.: Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. Trends in Cognitive Sciences 6, 505–510 (2002)
27. Piaget, J.: The origins of intelligence in childhood. International Universities Press (1952)
28. Prinz, W.: Ideo-motor action. In: Heuer, H., Sanders, A.F. (eds.) Perspectives on perception and action. Erlbaum, Hillsdale (1987)
29. Prinz, W.: A common coding approach to perception and action. In: Neumann, O., Prinz, W. (eds.) Relationships between perception and action, pp. 167–201. Springer, Berlin (1990)
30. Prinz, W.: Why don't we perceive our brain states? European Journal of Cognitive Psychology 4, 1–20 (1992)
31. Rumelhart, D.E., Hinton, G.E., McClelland, J.L.: A general framework for parallel distributed processing. In: Rumelhart, D.E., McClelland, J.L., The PDP Research Group (eds.) Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1, pp. 45–76. MIT Press, Cambridge (1986)
32. Schmidt, R.A.: A schema theory of discrete motor skill learning. Psychological Review 82, 225–260 (1975)
33. Stock, A., Stock, C.: A short history of ideo-motor action. Psychological Research 68, 176–188 (2004)
34. Thorndike, E.L.: Ideo-motor action. Psychological Review 20, 91–106 (1913)
35. Treisman, A.: The binding problem. Current Opinion in Neurobiology 6, 171–178 (1996)
36. Van der Velde, F., De Kamps, M.: From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation. Journal of Cognitive Neuroscience 13(4), 479–491 (2001)
37. Van der Velde, F., De Kamps, M., Van der Voort van der Kleij, G.: Clam: Closed-loop attention model for visual search. Neurocomputing 58-60, 607–612 (2004)
38. Wilson, H., Cowan, J.: Excitatory and inhibitory interactions in localized populations of model neurons. Biophysics Journal 12, 1–24 (1972)