RESEARCH REPORT

A computational model of perception and action for cognitive robotics

**Pascal Haazebroek · Saskia van Dantzig ·
Bernhard Hommel**

Abstract Robots are increasingly expected to perform tasks in complex environments. To this end, engineers provide them with processing architectures that are based on models of human information processing. In contrast to traditional models, where information processing is typically set up in stages (i.e., from perception to cognition to action), it is increasingly acknowledged by psychologists and robot engineers that perception and action are parts of an interactive and integrated process. In this paper, we present HiTEC, a novel computational (cognitive) model that allows for direct interaction between perception and action as well as for cognitive control, demonstrated by task-related attentional influences. Simulation results show that key behavioral studies can be readily replicated. Three processing aspects of HiTEC are stressed for their importance for cognitive robotics: (1) ideomotor learning of action control, (2) the influence of task context and attention on perception, action planning, and learning, and (3) the interaction between perception and action planning. Implications for the design of cognitive robotics are discussed.

Keywords Integrated processes · Perception–action interaction · Computational modeling · Cognitive robotics · Common coding · Ideomotor learning

P. Haazebroek (✉) · S. van Dantzig · B. Hommel
Cognitive Psychology Unit and Leiden Institute
for Brain and Cognition, Wassenaarseweg 52,
2333 AK Leiden, The Netherlands
e-mail: phaazebroek@fsw.leidenuniv.nl

Introduction

Robots are increasingly expected to autonomously fulfill a variety of tasks and duties in real-world environments that are constantly changing. In order to cope with these demands, robots cannot rely on predefined rules of behavior or fixed sets of perceivable objects or action routines. Rather, they need to be able to learn how to segregate and recognize novel objects (e.g., Kraft et al. 2008), how to perceive their own movements (e.g., Fitzpatrick and Metta 2003), and what actions can be performed on objects in order to achieve certain effects (e.g., Montesano et al. 2008).

In this ultimate robot engineering challenge, the nature of perception, action, and cognition plays an important role. Traditionally, these domains are assumed to reflect different stages of information processing (e.g., Donders 1868; Neisser 1967; Norman 1988): first, objects are perceived and recognized; subsequently, based on the current situation, task, and goal, the optimal action is determined; and finally, the selected action is prepared and executed. However, new robot architectures increasingly recognize the benefits of integration across these domains. Some roboticists have focused on creating (perceptually defined) anticipations that guide action selection and motor control (e.g., Hoffmann 2007; Ziemke et al. 2005). Others stress the importance of the acquisition and use of affordances (after Gibson 1979) in navigation (e.g., Uğur and Şahin 2010), action selection (e.g., Cos-Aguilera et al. 2004; Kraft et al. 2008), and imitation (e.g., Fitzpatrick and Metta 2003). Also, some approaches propose active perception strategies (e.g., Hoffmann 2007; Lacroix et al. 2006; Ognibene et al. 2008) that include epistemic sensing actions (e.g., eye movements) that actively perceive features necessary for object recognition or for the planning of further actions.

In our laboratory, we study perception and action in human performance. Findings in behavioral and neuro-cognitive studies have shed new light on the interaction between perception and action, indicating that these processes are not as separate and stage-like as has been presumed. First, features of perceived objects (such as location, orientation, and size) seem to influence actions directly and beyond cognitive control, as illustrated by stimulus-response compatibility phenomena, such as the Simon effect (Simon and Rudell 1967). This suggests that there is a direct route from perception to action that can bypass cognition. Second, in monkeys, neural substrates (i.e., so-called mirror neurons) have been discovered that are active both when the monkey performs a particular action and when it perceives the same action carried out by another monkey or human (Rizzolatti and Craighero 2004). This suggests that common representations exist for action planning and action perception. Finally, behavioral studies show that, in humans, action planning can actually influence object perception (Fagioli et al. 2007; Wykowska et al. 2009), suggesting that perceptual processes and action processes overlap in time.

In order to integrate these findings, we have developed a novel cognitive architecture that allows for the simulation of a variety of behavioral phenomena. We believe that our computational model includes processing aspects of human perception, action, and cognition that may be of special interest to designers of cognitive robots.

In this paper, we argue that in addition to representations that include both perceptual and action-related features (e.g., Wörgötter et al. 2009), perception, and action may also be intertwined with respect to their processes. First, we discuss the theoretical foundation of our work (the Theory of Event Coding: TEC) and describe our recently developed computational model, called HiTEC (Haazebroek et al. submitted). Then, we discuss a number of simulations of behavioral phenomena that illustrate the principles of the processing architecture underlying HiTEC. Finally, we discuss the wider implications of our approach for the design of cognitive robotics.

HiTEC

Theory of event coding

The theoretical basis of our approach is the Theory of Event Coding (TEC, Hommel et al. 2001), a general theoretical framework addressing how perceived events (i.e., stimuli) and produced events (i.e., actions) are cognitively represented and how their representations interact to generate perceptions and action plans. TEC claims that stimuli and actions are represented in the same way and by using the same "feature codes". These codes refer to the distal features of objects and events in the environment, such as shape, size, distance, and location, rather than to proximal features of the sensations elicited by stimuli (e.g., retinal location or auditory intensity). For example, a haptic sensation on the left hand and a visual stimulus on the left both activate the same distal code representing "left".

Feature codes can represent the properties of a stimulus in the environment just as well as the properties of a response—which, after all, is a perceivable stimulus event itself. This theoretical assumption is derived from ideomotor theory (James 1890; see Stock and Stock 2004, for a historical overview), which presumes that actions are cognitively represented in terms of their perceivable effects. According to the ideomotor principle, when one executes a particular action, the motor pattern is automatically associated to the perceptual input representing the action's effects (action effect learning: Elsner and Hommel 2001). Based on these action effect associations, people can subsequently plan and control (Hommel 2009) a motor action by anticipating its perceptual effects, that is, (re-)activate a motor pattern by intentionally (re-)activating the associated feature codes. Thus, stimuli and actions are represented in a common representational medium (Prinz 1990). Furthermore, stimulus perception and action planning are considered to be similar processes: both involve activating feature codes that represent external events.

Finally, TEC stresses the role of task context in stimulus and response coding. In particular, the responsiveness of feature codes to activation sources is modulated according to the task or goals at hand (the intentional weighting principle). For example, if the task is to grasp an object, features codes representing features relevant for grasping (such as the object's shape, size, location, and orientation) will be enhanced, while feature codes representing irrelevant features (such as the object's color or sound) will be attenuated.

In our work, we address how these principles may be computationally realized. To this end, we have developed a computational model, called HiTEC, and tested its performance against empirical data from human studies on various types of perception–action interactions (Haazebroek et al. submitted). In this paper, two simulations are described, illustrating the aspects that are of particular relevance for information processing in cognitive robotics.

HiTEC's structure and representations

HiTEC is implemented as a connectionist network model that uses the basic building blocks of parallel distributed processing (PDP; e.g., Rumelhart et al. 1986). In a PDP model, processing occurs through the interactions of a number of interconnected elements called units. Units may

be organized into higher order structures, called modules. Each unit has an activation value indicating local activity. Processing occurs by propagating activity through the network, that is, by propagating activation from one unit to the other, via weighted connections. When a connection between two units is positively weighted, the connection is excitatory and the units will increase each other's activation. When the connection is negatively weighted, it is inhibitory and the units will reduce each other's activation. Processing starts when one or more units receive some sort of external input. Gradually, unit activations will change and propagate through the network while interactions between units control the flow of processing. Some units are designated output units. When the activation of any one of these units reaches a certain threshold, the network is said to produce the corresponding output.

In HiTEC, the elementary units are codes that may be connected and are contained within maps (HiTEC's modules). Codes within the same map compete for activation by means of lateral inhibitory connections. As illustrated in Fig. 1, maps are organized into three levels: the sensory-motor level, the feature level, and the task level. Each level will now be discussed in more detail.

Sensory-motor level

The primate brain encodes perceived objects in a distributed fashion; different features are processed and represented across different cortical maps (e.g., DeYoe and Van Essen 1988). In HiTEC, different perceptual modalities (e.g., visual, auditory, tactile, proprioceptive) 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 contains a number of sensory codes that are responsive to specific sensory features (e.g., a specific color or a specific pitch). Note that Fig. 1 shows only the sensory maps that are relevant for modeling the Simon effect (Simulation 2 in the Simulations section): auditory pitch, auditory location, and haptic location. However, other specific instances of the model may include other sensory maps as well (e.g., visual maps).

The sensory-motor level also contains motor codes, referring to more or less specific movements (e.g., the muscle contractions that produce the movement of the hand pressing a certain key). Although motor codes could also be organized in multiple maps, in the present version of HiTEC, we consider only one basic motor map with a set of motor codes.

Feature level

TEC's notion of "feature codes" is captured by codes that are connected to and thus grounded in both sensory codes and motor codes. Crucially, the same (distal) feature code (e.g., "left") can be connected to multiple sensory codes (e.g., "left haptic location" and "left auditory location"). Thus, information from different sensory modalities and dimensions is combined in one feature code representation. Although feature codes are considered to arise from experience, in the present HiTEC simulations, we assume the existence of a set of feature codes (and their connections to sensory codes) to bootstrap the process of extracting sensorimotor regularities in interactions with the environment.

Fig. 1 HiTEC architecture as used for simulation of the Simon effect (see "Simulations section"). Codes reside in maps on different levels and are connected by excitatory associations. *Solid lines* denote fixed weights, *dashed lines* are connections with learned weights. Sensory codes receive modulated excitation from feature codes, denoted by the *open arrows*
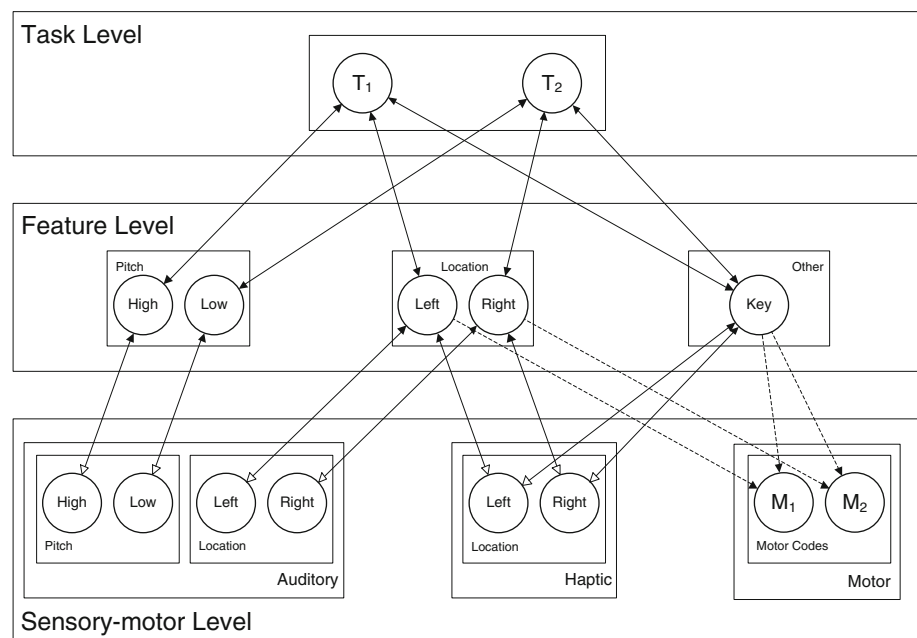
Task level

The task level contains generic task codes that reflect alternative ways to wire and weigh existing representations to prepare for and carry out a particular task. Task codes connect to both the feature codes that represent stimuli and the feature codes that represent responses, in correspondence with the current task context.

Associations

In HiTEC, codes are associated. Some are considered innate or reflecting prior experience (depicted as solid lines in Fig. 1), others are learned during the simulation (depicted as dashed lines in Fig. 1). This will be elaborated in the next subsection.

HiTEC's processes

In general, codes can be stimulated, which results in an increase in their activation level. Gradually, activation will flow toward other codes through the connections. Note that connections are bidirectional (except for the learned feature code—motor code associations), which results in activation flowing back and forth between sensory codes, feature codes, and task codes and activation flowing toward motor codes. In addition, codes within the same map inhibit each other. Together, this results in a global competition mechanism in which all codes participate from the first processing cycle to the last.

Ideomotor learning

Associations between feature codes and motor codes are explicitly learned as follows. A random motor code is activated (comparable to the spontaneous "motor babbling" behavior of newborns) first. This leads to a change in the environment (e.g., the left hand suddenly touches an object) that is registered by sensory codes. Activation propagates from sensory codes toward feature codes. Subsequently, the system forms associations between the active feature codes and the active motor code. The weight change of these associations depends on the level of activation of both the motor code and the feature codes during learning.

Action planning

Once associations between motor codes and feature codes exist, they can be used to select and plan actions. Planning an action is realized by activating the feature codes that correspond to its perceptual effects and by propagating their activation toward the associated motor codes. Initially, multiple motor codes may become active as they typically fan out

associations to multiple feature codes. However, some motor codes will have more associated features and some of the associations between motor codes and feature codes may be stronger than others. In time, the network converges toward a state where only one motor code is strongly activated, which leads to the selection of that motor action.

Task preparation

In behavioral experiments, participants typically receive a verbal instruction of the task. In HiTEC, a verbal task instruction is assumed to directly activate the respective feature codes. The cognitive system connects these feature codes to task codes. When the model receives several instructions to respond differently to various stimuli, different task codes are recruited and maintained for the various options. Due to the mutual inhibitory links between these task codes, they will compete with each other during the task. Currently, the associations between feature codes and task codes are set by hand.

Responding to stimuli

When a stimulus in an experimental trial is presented, its sensory features will activate a set of feature codes allowing activation to propagate toward one or more task codes, which were already associated during task preparation. Competition takes place between feature codes, between task codes, and between motor codes, simultaneously. Once any one of the motor codes is activated strongly enough, it leads to the execution of the respective motor response to the presented stimulus. In our simulations, this marks the end of a trial. In general, the passing of activation between codes along their connections is iterated for a number of cycles, which allows for the simulation of reaction time (i.e., number of cycles from stimulus onset to response selection) until the activation level of any one of the motor code reaches a set threshold value.

Neural network implementation

All HiTEC codes—sensory codes, motor codes, task codes, and feature codes alike—are implemented as generic units with an activation value bound between 0.0 and 1.0. Stimulus presentation is simulated by setting the external input values for the sensory codes. In addition, codes receive input from other, connected codes. In sum, their activation level is updated each processing cycle according to the following function:

$$A(t+1) = (1-d) \times A(t) + (Ext + Exc + Bg) \times (1 - A(t)) + Inh \times A(t)$$

Here, A(t) denotes the activation of node at time t, d is a decay term, Ext is its external input (sensory codes only),

Exc is its excitatory input (bottom up and top down), *Bg* is additive background noise, and *Inh* is its inhibitory input. Note that we first compute all input values to all codes and then update their respective activation values, resulting in (simulated) synchronous updating. Ideomotor weights (i.e., weights of connections between feature codes and motor codes) are acquired during learning trials, using the following Hebbian learning rule:

$$w_{jk}(t+1) = (1-d) \times w_{jk}(t) + LR \times Act_j(t) \times Act_k(t) \times (1 - w_{jk}(t))$$

Here, w_{ij} denotes the weight between nodes j and k at time t, d is a decay term, LR is the learning rate, and Act_j and Act_k refer to the activation levels at time t of nodes j and k, respectively.

Note that we register a stimulus (or action effect) by simply providing input values to the relevant sensory code(s). In a real robotic setup, these sensory codes could be grounded in the environment, either by defining this a priori or by generating the sensory codes by means of unsupervised clustering techniques (e.g., the category formation phase in Montesano et al. 2008). Also note that, in principle, our model is not restricted to any number or type of sensory codes, as long as one can find a reasonable way to ground their stimulation in actual (robotic) sensor values. In simulations, we use specific instances of the model that include only those codes we need for the simulation at hand. Nothing, however, prohibits us from including other codes as well. The fact that only those codes that are connected to higher level feature and task codes (resulting from the task internalization) receive top-down enhancement makes that only a selection of available codes matter for further processing and other sensory stimulation is to be ignored (unless strongly salient). Also note that the time dimension is actually of value to us. Rather than computing a perception–action mapping or function as fast as possible—as is common in robotics algorithms—taking more time for a computation actually reflects the (cognitive or perception–action related) effort needed for the task at hand. Consequently, apart from accuracy scores, we also obtain reaction times that vary as a function of the difficulty of the task. For further computational details, we refer to Haazebroek et al. (submitted).

Simulations

The HiTEC computational model has been tested against a variety of behavioral phenomena. Three processing aspects of the model are of particular interest: (1) ideomotor learning of action control, (2) the influence of task context and at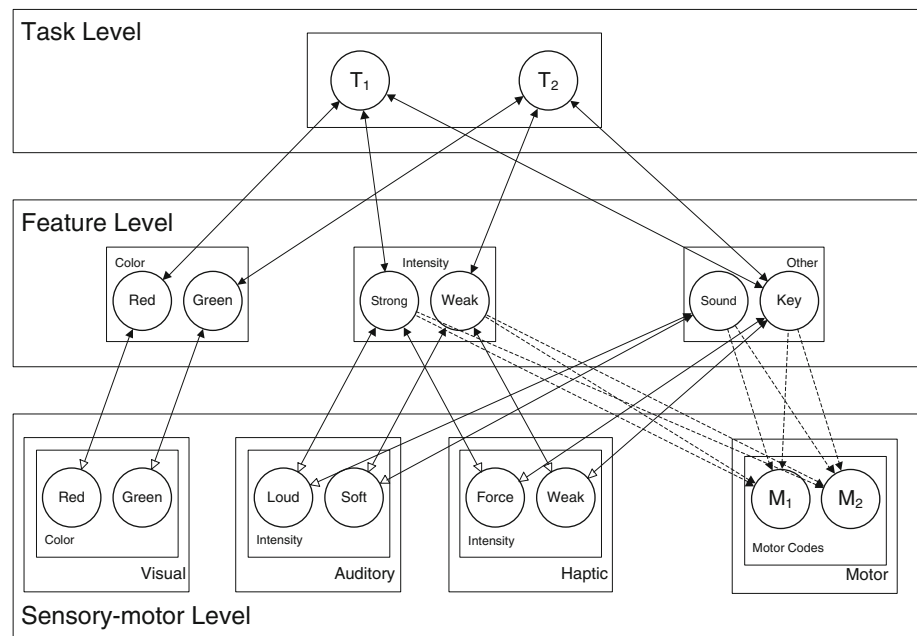tention on both perception and action planning, and (3) the interaction between perception and action planning. For the purpose of this paper, these processing aspects are illustrated by simulating two behavioral studies: a study on response-action effect compatibility, conducted by Kunde et al. (2004) and the Simon task (Simon and Rudell 1967). We now elaborate on these paradigms and discuss the results of the simulations in HiTEC.

Simulation 1

In earlier studies (e.g., Elsner and Hommel 2001), it has been shown that people automatically learn associations between motor actions and their perceptual effects. An experiment by Kunde et al. (2004) demonstrates that performance is also affected by the compatibility between responses and novel (auditory) action effects. In this experiment, participants had to respond to the color of a visual stimulus by pressing a key forcefully or softly. The key presses were immediately followed by a loud or soft tone. For one group of participants, responses were followed by a compatible action effect; the loudness of the tone matched the response force (e.g., a loud tone appeared after a forceful key press). In the other group of participants, the relationship between actions and action effects was incompatible (e.g., a soft tone appeared after a forceful key press). If the intensity of an action (e.g., a forceful response) was compatible with the intensity of the action effect (e.g., a loud tone), then responses were faster than if the intensity of the action and its effect were incompatible. Given that the tones did not appear before the responses were executed, this observation suggests that the novel, just acquired action effects were anticipated and considered in the response selection process.

In HiTEC, this experiment is simulated using the model as depicted in Fig. 2. Here, visual colors, auditory intensity, and haptic intensity are coded by their respective sensory codes. At the feature level, generic intensity feature codes exist that code for both auditory and haptic intensity. The motor codes refer to the soft and forceful key presses, respectively. The bindings between feature codes and task codes follow the task instructions: red is to be responded to by a strong key press, green by a weak key press. As an equivalent of the behavioral study, we allow the model to learn the relationship between key presses and their perceptual effects. This is realized by randomly activating one of the two motor codes (i.e., pressing the key softly or forcefully), registering their sensory effects (i.e., both the key press and the tone) by activating the respective sensory codes, and propagating activation toward the feature codes. As a result, the active motor code and the active feature codes become (more strongly) associated. Thus, ideomotor learning takes place; actions are associated with their effects.

Importantly, trials with incompatible action effects result in simultaneous activation of both "strong" and "weak" feature codes. Moreover, these feature codes inhibit each other. As a result, associations learned between feature codes and motor codes are weaker (and less specific) than in trials with compatible action effects. As these associations (co-)determine the processing speed during experimental trials, group differences in reaction time arise: 24.16 cycles (SD = 0.20) for the compatible group and 30.64 cycles (SD = 4.10) for the incompatible group. These results fit well with available behavioral data as shown in Fig. 3. Thus, regular and compatible action effects allow for faster anticipation of a motor action resulting in faster responses to stimuli in the environment than when action effects are incompatible.

Crucially, participants in behavioral studies typically receive a task instruction before performing practice trials. This is simulated by internalizing the task instruction as feature code—task code connections already before performing the ideomotor learning. As a consequence, activation also propagates from the haptic intensity sensory codes to the "Key" feature code to the task codes—and back—during ideomotor learning, resulting in a top-down enhancement of the haptic intensity sensory codes as compared to the auditory intensity sensory codes. While both haptic intensity and auditory intensity sensory codes receive equal external stimulation from the environment, haptic intensity comes to determine the learned connection weights, where auditory intensity only moderates this process. Thus, the task instruction not only specifies the actual S-R mappings that reflect the appropriate responses to the relevant stimulus features, it also configures the

model in such a way that "most attention is paid" to those action effect features in the environment that matter for the current task. Indeed, additional action effects (i.e., the auditory intensity in this task) do not escape processing completely (hence the compatibility effect), but their influence is strongly limited as compared to the main action effect dimension (i.e., the haptic intensity). This illustrates the role of task context and attention in ideomotor learning of action control.

Simulation 2

The canonical example of stimulus-response compatibility effects is the Simon task. In this task, the participant performs manual, spatially defined responses (e.g., pressing a left or right key) to a nonspatial feature of a stimulus (e.g., the pitch of a tone). Importantly, the location of the stimulus varies randomly. Even though stimulus location is irrelevant for the response choice, performance is facilitated when stimulus location corresponds spatially to the correct response. Conversely, performance is impaired when stimulus location corresponds spatially to the other, not to be chosen response.

The Simon task is simulated in HiTEC using the model as depicted in Fig. 1. There are sensory codes for auditory pitch levels, auditory locations, and haptic locations. The motor codes refer to the left key press and the right key press motoric responses, respectively. At the feature level, generic pitch feature codes as well as generic location codes are included, as is a general "Key" code that allows the system to code for all key-related features in a simple way.

Fig. 3 Simulation results (*left panel*) compared with behavioral data (*right panel*, adopted from Kunde et al., 2004)
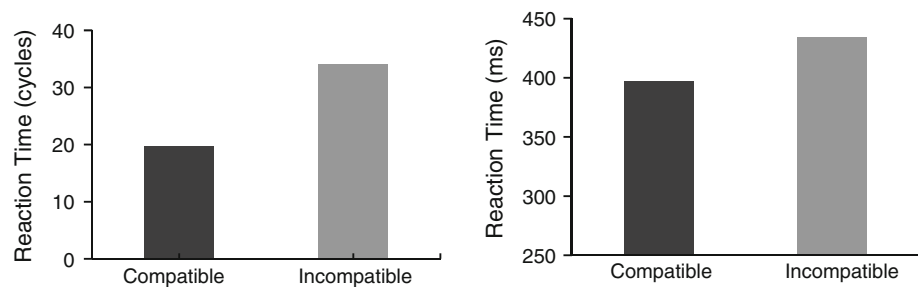
Fig. 4 Results of simulation of the Simon task (*left panel*) compared with behavioral data (*right panel*, adopted from Simon and Rudell 1967)
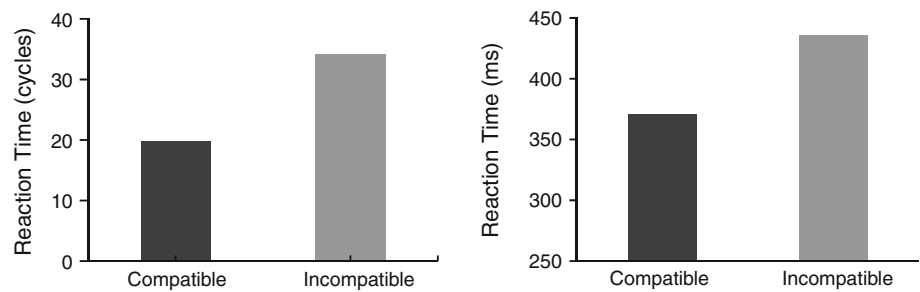
Crucially, the "Left" and "Right" feature codes are used both for encoding (egocentric) stimulus location (i.e., they are connected to auditory location sensory codes) and for encoding (egocentric) response location (i.e., they are connected to haptic location sensory codes). When a tone stimulus is presented, auditory sensory codes are activated, activation propagates gradually toward pitch and location feature codes, toward task codes and (again) toward location feature codes. Over time, the motor codes also become activated until one of the motor codes reaches a threshold level and is executed. Simulation results thus yield a clear compatibility effect. On average, the compatible trials require 19.79 cycles (SD: 0.18), neutral trials 25.28 (SD: 0.23) cycles, and incompatible trials 34.04 (SD: 0.73) cycles. These results fit well with the available behavioral data (Simon and Rudell 1967) as is shown in Fig. 4.

To understand the source of the compatibility effect in this model, consider the following compatible trial: a high tone is presented on the left. This activates the auditory sensory codes S_{high} and S_{left}. Activation propagates to the feature codes F_{high} and F_{left} and to task code T_1. Because the task is to respond to high tones with a left key press, activation flows from T_1 to F_{left} and from there to M_1 (left response). As soon as the motor activation reaches a threshold, the action is executed (i.e., the left key is pressed). Now, consider an incompatible trial: a high tone is presented on the right. This activates the auditory sensory codes S_{high} and S_{right}. Activation propagates to the feature codes F_{high} and F_{right} and to task code T_1. Activation flows from T_1 to F_{left}. Because both F_{left} and F_{right} are activated,

activation is propagated to both motor codes. As a result, competition arises at different levels of the model, which results in a longer response time.

Thus, because common feature codes are used to represent both stimulus features and response features, stimulus-response compatibility effects are bound to occur. In the case of this simulation of the Simon effect, although the location of the stimulus is irrelevant for the task, location is relevant for the anticipation (thus planning) of the appropriate response. Now, the model is, in a sense, forced to also process the stimulus location and—automatically—bias the planning of the action response, generating the Simon effect. Thus, again ideomotor learning is influenced by the current task. In addition, automatic interaction between perception and action causes compatibility effects. Still, even this automatic interaction arises from the task that defines the stimulus features and response (effect) features that are to be used in planning a response to the presented stimulus.

Discussion

In this paper, we argue that perception and action are not only intertwined with regard to their representations, but may also be intertwined with respect to their processes, in both humans and artificial cognitive systems. To demonstrate this, we have presented a computational model, HiTEC, which is able to replicate key findings in psychological research (Haazebroek et al. submitted). Three

processing aspects of the model may be of particular interest for cognitive robotics: ideomotor learning of action control, the role of task context and attention, and the interaction between perception and action. These aspects, as well as implications for anticipation and affordances—two important themes in cognitive robotics—are discussed.

Ideomotor learning of action control

The first aspect concerns the ideomotor learning of action control. When working with robots in real-world environments, it is quite common to assume a predefined set of (re)actions and let the robot learn whether these actions can be applied to certain objects (e.g., Cos-Aguilera et al. 2004; Kraft et al. 2008). However, it is often hard to define a priori exactly which actions can be performed by the robot given the characteristics of the robot body and the physical environment. Moreover, the same robot body can have different action capabilities in different (or changing) environments. Thus, it makes sense not to pre-wire the actions in the system, but let them be learned by experience. In our model, we explicitly do not include a known set of movements. Instead, motor actions are labeled "M1" and "M2" on purpose; only by means of sensorimotor experience, associations are learned between feature codes and motor codes, allowing the model to represent its own actions in terms of the perceived action effects. The model can subsequently plan a motor action by anticipating (i.e., activating the feature codes corresponding to) these perceptual effects (see also the subsection on anticipation). The theme of ideomotor learning is already quite popular in the robotics literature (for an overview, see Pezzulo et al. 2006). Note that in our current implementation, we assume the existence of distal feature codes and let the model learn connections between these feature codes and motor codes, rather than between sensory codes and motor codes directly. We envision that this generalization renders the acquired ideomotor associations applicable to a variety of circumstances (i.e., generalizing over objects and over actions). In addition, ideomotor learning is modulated by the current task and attentional resources, as described in the next subsection.

Task context and attention

The second aspect is the influence of task context and attention on both perception and action. As is demonstrated in Simulation 1, internalization of the task instruction not only specifies the stimulus-response mappings that determine the perception-cognition-action information flow from stimulus to response, but also influences the interactions between perception and action (see next subsection). Codes that are relevant for stimulus perception and/or

action planning are top-down enhanced. Different task contexts may weigh different dimensions and features differently and consequently speed up the stimulus-response translation process in a way that suits the task best. Attention is studied quite extensively in cognitive robotics and is key in active perception strategies (e.g., Ognibene et al. 2008).

In HiTEC, feature codes belonging to objects can be enhanced in activation (in other words: attended to), as can be features belonging to action effects (in other words: intended or effect anticipation). Thus, as demonstrated by Simulation 1, task context can make the system "pay attention" to those action effects that are important to the task while attenuating the influence of irrelevant additional effects. This processing aspect is crucial when learning (both in sensorimotor and ideomotor learning) in a complex environment. As cognitive robots are typically able to track a wealth of sensor readings, it can be hard to determine which action effects are particularly relevant for learning (see also the frame problem in artificial intelligence, Russell and Norvig 1995; Dennett 1984). The selective attention for elements of the important action effects still relies on the context such as the task at hand or on a system that uses internal drives (e.g., Cos-Aguilera et al. 2004), but the weighting of action effect features may present an elegant way for a robot system to specifically monitor its anticipated action results.

Also, actions are selected based on perceived object features—following the task level connections—and the attention for specific features or feature dimensions applied to both object perception and action planning. As a consequence, the model does not select actions for a particular object in a reflex-like manner, but it takes the current task into account, both in object perception and response selection (see also Hommel 2000).

It is important to note that our conception of "task" is extremely simplified and only relates to the current stimulus to response translation rules, neglecting the human capacity of planning action sequences or even performing multiple tasks simultaneously (e.g., as modeled in EPIC, Kieras and Meyer 1997). However, HiTECs task could be part of a broader definition of task, including sequences of actions and criteria for determining whether a task has ended (see also Pezzulo et al. 2006 for a conceptual analysis of ideomotor learning and a task-oriented model such as TOTE).

Interaction between perception and action

The third aspect is the interaction between perception and action planning as a result of the common coding principle (Hommel et al. 2001). Feature codes that are used to cognitively represent stimulus features (e.g., object

location, intensity) are also used to represent action features. As a result, stimulus-response compatibility effects can arise, such as the Simon effect that we replicated in Simulation 2: when a feature code activated by the stimulus is also part of the features belonging to the correct response, planning this response is facilitated, yielding faster reactions. If, on the other hand, the feature code activated by the stimulus is part of the incorrect response, this increases the competition between the alternative responses, resulting in slower reactions. It may appear that this effect is due to information processing from stimulus to response only. However, following the HiTEC logic, the stimulus location only has an effect because location is relevant for the action. Thus, action planning influences object perception. Moreover, while attention plays a role in "tagging" codes to be of importance (whether it is for stimulus perception or for action planning), the actual information flow from stimulus to response (effect) is rather automatic, in a sense bypassing the "cognitive" stage. In robotics, one could envision that perceiving an object shape already enhances the planning of a roughly equivalent hand shape. In our model, this is done on a distal level. That is, a ball-park type of "equivalent" shape is anticipated. During a reaching movement, the precise hand shape can be adjusted to the precise shape of the object to be grasped. Of course, how this is realized exactly should be subject to further research. Still, the automatic translation from perceived object features to anticipated action features seems a promising characteristic with respect to the notion of affordance (see last subsection).

Anticipation

Numerous cognitive robotics projects include a notion of "anticipation" (see Pezzulo et al. 2008 for an overview). This is crucial, because in action planning and control it seems difficult to evaluate all potential variations in advance, as real data can vary a lot and the behavior of the environment is not always completely controlled by the robot. Thus, it seems intuitive to form anticipations on a distal level. In HiTEC, we adhere to an explicit notion of anticipation using distal feature codes. Activating these codes not only helps in action selection or planning, rather it is how actions are selected and planned in HiTEC. Following the ideomotor principle, the anticipated action effect activates the associated motor action resulting in actual action execution. However, how, the action subsequently unfolds is not explicitly modeled. By comparing the anticipated effects with the actually perceived effects, the system can determine whether the action was successful. If there is a discrepancy between the anticipated and perceived action effect, the model can update its representations to learn new action effects or fine-tune action

control (see for more details Haazebroek and Hommel 2009).

Finally, as described in previous subsections, the system focuses attention on those features that are important for the current task context, both in object perception and in action effect perception. Thus, the task context determines which effect features are especially monitored. This process may present opportunities to tackle the frame problem in artificial intelligence (Russell and Norvig 1995; Dennett 1984).

Affordances

Roboticists have started to embrace the notion of affordance (after Gibson 1979) in their robot architectures. For example, Montesano et al. (2008) define affordances as links between objects, actions, and effects. It is commonly assumed that such links are acquired during experience. Typical setups consist of an exploration phase in which actions are executed randomly and action success is determined. Then, object features are correlated with actions and their success. Finally, for each action, it is determined which object features are good indications for their successful execution yielding a set of affordances in terms of stimulus features—motor actions (S-R) reflexes.

In our model, the notion of affordance is effectively realized by allowing for automatic translation of perceptual object features (e.g., object shape) to action by means of overlap with anticipated action effect features (e.g., hand shape). In this sense, feature codes are representations of regularities encountered in sensorimotor experience. By focusing attention on certain action plan features, these dimensions also become enhanced in object perception. As a consequence, these sensory features are processed more strongly than others. Thus, rather than defining affordances as successful, yet arbitrary, S-R reflexes, we define an affordance in terms of intrinsic overlap between stimulus features and action effect features as encountered in sensorimotor experience. Indeed, by having common codes, perceiving objects fundamentally implies anticipating intrinsically related action plan features. Because of the ideomotor links, activating features shared by objects and actions, the system is easily biased to plan and execute appropriate motor actions. Crucially, the task determines which (object and action) features are relevant and therefore which affordances apply.

Conclusion

We have shown how the HiTEC model can readily replicate key findings from the perception–action literature using simulations. These simulations demonstrate three

main processing aspects that seem relevant for cognitive robots as they intimately relate to crucial themes as anticipation and affordances.

It is clear that HiTEC is still limited to simulations of basic, yet fundamental phenomena in the perception–action domain. Our main objective was to develop an architecture that gives an integrated processing account of the interaction between perception and action. However, we envision that the model could be extended with a variety of capabilities, such as higher fidelity perception and motor action, as well as episodic and semantic memory capacities. This way, HiTEC could become a more mature cognitive architecture (Byrne 2008).

Other extensions may include the processing of affective information. In reinforcement learning approaches, affective information is usually treated as additional information that co-defines the desirability of a state (i.e., as a "reward") or action alternative (i.e., as part of its "value" or "utility"). By weighting action alternatives with this information (see also the notion of somatic markers, Damasio 1994), some can turn out to be more desirable than others, which can aid the process of decision making (e.g., Broekens and Haazebroek 2007). In psychological research, studies have shown affective stimulus-response compatibility effects (e.g., Chen and Bargh 1999; van Dantzig et al. 2008). Participants are typically faster to respond to positive stimuli (e.g., the word "love", a picture of a smiling face) when they perform an approach movement than an avoidance movement. Conversely, they are faster to respond to negative stimuli (e.g., the word "war", a picture of a spider) when performing an avoidance movement than an approach movement. These findings are taken as evidence that affective stimuli automatically activate action tendencies related to approach and avoidance (e.g., Chen and Bargh 1999). Elsewhere (Haazebroek et al. 2009), we have shown how HiTEC can already account for such affective stimulus-response compatibility effects.

Still, HiTEC has not yet been implemented on an actual robot platform, which would require substantial implementation effort concerning the grounding of sensory input and motor output. However, it would also provide a wealth of opportunities to study and test HiTEC's processing aspects in real-world scenarios.

References

Broekens J, Haazebroek P (2007) Emotion and reinforcement: affective facial expressions facilitate robot learning. In: Proceedings of the IJCAI workshop on AI for human computing (AI4HC'07), pp 47–54

Byrne MD (2008) Cognitive architecture. In: Jacko JA, Sears A (eds) Human-computer interaction handbook. Erlbaum, Mahwah, pp 97–117

Chen S, Bargh JA (1999) Consequences of automatic evaluation: immediate behavior predispositions to approach or avoid the stimulus. Pers Soc Psychol Bull 25:215–224

Cos-Aguilera I, Hayes G, Canamero L (2004) Using a SOFM to learn object affordances. In: Proceedings of the 5th workshop of physical agents (WAF 04)

Damasio AR (1994) Descartes' error. G. P. Putnam's Sons, New York

Dennett DC (1984) Cognitive wheels: the frame problem of AI. In: Hookaway C (ed) Minds, machines and evolution. Cambridge University Press, Cambridge, pp 129–151

DeYoe EA, Van Essen DC (1988) Concurrent processing streams in monkey visual cortex. Trend Neurosci 11:219–226

Donders FC (1868) Over de snelheid van psychische processen. Onderzoekingen, gedaan in het physiologisch laboratorium der Utrechtsche hoogeschool, 2. Reeks 2:92–120

Elsner B, Hommel B (2001) Effect anticipation and action control. J Exp Psychol Hum Percept Perform 27:229–240

Fagioli S, Hommel B, Schubotz RI (2007) Intentional control of attention: action planning primes action-related stimulus dimensions. Psychol Res 71:22–29

Fitzpatrick P, Metta G (2003) Grounding vision through experimental manipulation. Philos Transact A Math Phys Eng Sci 361:2165–2185

Gibson JJ (1979) The ecological approach to visual perception. Houghton Mifflin, Boston

Haazebroek P, Hommel B (2009) Anticipative control of voluntary action: towards a computational model. In Pezzulo G, Butz MV, Sigaud O, Baldassarre G (eds) Anticipatory behavior in adaptive learning systems: from psychological theories to artificial cognitive systems. Lecture Notes in Artifical Intelligence, vol 5499, pp 31–47

Haazebroek P, van Dantzig S, Hommel B (2009). Towards a computational account of context mediated affective stimulus-response translation. In: Proceedings of the 31st annual conference of the cognitive science society. Cognitive Science Society, Austin, TX

Haazebroek P, Raffone A, Hommel B (submitted). HiTEC: a computation model of the interaction between perception and action planning

Hoffmann H (2007) Perception through visuomotor anticipation in a mobile robot. Neural Netw 20:22–33

Hommel B (2000) The prepared reflex: automaticity and control in stimulus-response translation. In: Monsell S, Driver J (eds) Control of cognitive processes: attention and performance XVIII. MIT Press, Cambridge, pp 247–273

Hommel B (2009) Action control according to TEC (theory of event coding). Psychol Res 73:512–526

Hommel B, Müsseler J, Aschersleben G, Prinz W (2001) The theory of event coding (TEC): a framework for perception and action planning. Behav Brain Sci 24:849–937

James W (1890) The principles of psychology. Dover Publications, New York

Kieras DE, Meyer DE (1997) An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. Hum Comput Interact 12:391–438

Kraft D, Pugeault N, Baseski E, Popovic M, Kragic D, Kalkan S et al. (2008) Birth of the object: detection of objectness and extraction

of object shape through object- action complexes. Int J HR 5:247–265

Kunde W, Koch I, Hoffmann J (2004) Anticipated action effects affect the selection, initiation, and execution of actions. Q J Exp Psychol A 57:87–106

Lacroix JPW, Postma EO, Hommel B, Haazebroek P (2006) NIM as a brain for a humanoid robot. In: Proceedings of the toward cognitive humanoid robots workshop at the IEEE-RAS international conference on humanoid robots 2006. Genoa, Italy

Montesano L, Lopes M, Bernardino A, Santos-Victor J (2008) Learning object affordances: from sensory motor coordination to imitation. IEEE Trans Robot 24:15–26

Neisser U (1967) Cognitive psychology appleton-century-crofts. New York

Norman DA (1988) The design of everyday things. Basic Book, New York

Ognibene D, Balkenius C, Baldasarre G (2008) Integrating epistemic action (active vision) and pragmatic action (reaching): a neural architecture for camera-arm robots. In: Minoru A, Hallam JCT, Mayer J-A, Tani J (eds) From animals to animats 10

Pezzulo G, Baldassarre G, Butz MV, Castelfranchi C, Hoffmann J (2006) An analysis of the ideomotor principle and TOTE. In: Butz MV, Sigaud O, Pezzulo G, Baldassarre G (eds) Anticipatory behavior in adaptive learning systems: advances in anticipatory processing. Springer LNAI 4520, pp 73–93

Pezzulo G, Butz MV, Castelfranchi C (2008) The anticipatory approach: definitions and taxonomies. In: Pezzulo G, Butz MV, Castelfranchi C, Falcone R (eds) The challenge of anticipation: a unifying framework for the analysis and design of artificial cognitive systems. Springer LNAI 5225, pp 23–43

Prinz W (1990) A common coding approach to perception and action. In: Neumann O, Prinz W (eds) Relationships between perception and action. Springer Verlag, Berlin, pp 167–201

Rizzolatti G, Craighero L (2004) The mirror-neuron system. Annu Rev Neurosci 27:169–192

Rumelhart DE, Hinton GE, McClelland JL (1986) A general framework for parallel distributed processing. In: Rumelhart DE, McClelland JL (eds) Parallel distributed processing: explorations in the microstructure of cognition. MIT Press, Cambridge, MA, pp 45–76

Russell S, Norvig P (1995) Artificial intelligence: a modern approach. Prentice Hall, NJ

Simon JR, Rudell AP (1967) Auditory S-R compatibility: the effect of an irrelevant cue on information processing. J Appl Psychol 51:300–304

Stock A, Stock C (2004) A short history of ideo-motor action. Psychol Res 68:176–188

Uğur E, Şahin E (2010) Traversability: a case study for learning and perceiving affordances in robots. Adaptive Behavior 18:258–284

van Dantzig S, Pecher D, Zwaan RA (2008) Approach and avoidance as action effects. Q J Exp Psychol 61:1298–1306

Wörgötter F, Agostini A, Krüger N, Shylo N, Porr B (2009) Cognitive agents—a procedural perspective relying on the predictability of Object-Action-Complexes (OACs). Rob Auton Syst 57:420–432

Wykowska A, Schubö A, Hommel B (2009) How you move is what you see: action planning biases selection in visual search. J Exp Psychol Hum Percept Perform 35:1755–1769

Ziemke T, Jirenhed D-A, Hesslow G (2005) Internal simulation of perception: a minimal neuro-robotic model. Neurocomputing 68:85–104