Information Sciences 186 (2012) 1–19

Contents lists available at SciVerse ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

PASAR: An integrated model of prediction, anticipation, sensation, attention and response for artificial sensorimotor systems

Zenon Mathews [a,*], Sergi Bermúdez i Badia [a], Paul F.M.J. Verschure [a,b]

[a] SPECS, Technology Department, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain
[b] Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona , Spain

ARTICLE INFO

Article history:
Received 4 January 2010
Received in revised form 6 July 2011
Accepted 27 September 2011
Available online 14 October 2011

Keywords:
Autonomous cognitive agents
Attention
Limited resources
Data integration
Anticipatory behavior

ABSTRACT

A wide range of neuroscientific studies suggest the existence of cognitive mechanisms like attention, prediction, anticipation and strong vertical interactions between different hierarchical layers of the brain while performing complex tasks. Despite advances in both cognitive brain research and in the development of brain-inspired artificial cognitive systems, the interplay of these key ingredients of cognition remain largely elusive and unquantified in complex real-world tasks. Furthermore, it has not yet been demonstrated how a self-contained hierarchical cognitive system acting under limited resource constraints can quantifiably benefit from the incorporation of top–down and bottom–up attentional mechanisms. In this context, an open fundamental question is how a data association mechanism can integrate bottom–up sensory information and top–down knowledge. Here, building on the Distributed Adaptive Control (DAC) architecture, we propose a single framework for integrating these different components of cognition and demonstrate the framework's performance in solving real-world and simulated robot tasks. Using the model we quantify the interactions between prediction, anticipation, attention and memory. Our results support the strength of a complete system that incorporates attention, prediction and anticipation mechanisms compared to incomplete systems for real-world and complex tasks. We unveil the relevance of transient memory that underlines the utility of the above mechanisms for intelligent knowledge management in artificial sensorimotor systems. These findings provide concrete predictions for physiological and psychophysical experiments to validate our model in biological cognitive systems.

1. Introduction

Biological systems excel in solving simple and complex tasks in dynamic environments facing limited resource constraints. Despite many decades of research in cognitive science and robotics, artificial systems are still far removed from the efficiency and robustness of natural behaving systems. It is also for this reason that the interest in biomimetics has been growing based on the belief that understanding biological systems can enhance our ability to design and deploy new generations of artificial sensori-motor systems. A self-contained real-world cognitive system has to deal with both external and internal limited resource constraints. External constraints impinge on the survival and overall integrity of the behaving system. Whereas internal constraints result from the physical instantiation of the system. In general, these internal constraints occur in all of the three stages of the perception-cognition-action cycle: (1) perception is limited due to a limited number of sensors and restrictions to their spatio-temporal resolution, (2) cognition is restricted due to limited memory capacity and

* Corresponding author. Tel.: +34 93 542 22 5 01; fax: +34 93 542 22 02.
E-mail address: zenon.mathews@upf.edu (Z. Mathews).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2011.09.042

computational resources and (3) action is limited by the degrees of freedom the system can control and their dependencies. Hence, given the general nature of these internal constraints any cognitive real-world agent is forced to use its resources selectively. To achieve this selectivity at different levels of the perception-cognition-action cycle, the biological brain employs a mechanism generally known as *attention* [26,32,33,58].

The role of bottom–up and top–down information flow in the brain has been of great interest for understanding attentional mechanisms involved in biological cognition [22,33,35]. Bottom–up information refers to the instantaneous information about the environment available to a real-world cognitive system through its sensors, whereas top–down information refers to the use of knowledge previously acquired by the system about itself, the environment and the task at hand. Top–down information processing has been strongly tied to the notion of prediction and its role in increasing the resistance to noise [15,17,63]. Some theories, like the so called reverse hierarchy model propose a strict top–down nature of some aspects of cognition (e.g. perceptual learning [3]), while others support the notion of parallel bottom–up and top–down processing streams [32,33,69].

Despite the elaborate ongoing discussion of bottom–up and top–down attentional processes, the mechanisms behind the integration of these two distinct processes of attention in the biological brain is not yet clearly known. Traditionally, the formation of individual memory items from bottom–up sensory input and the top–down search of particular memory items have been treated separately [33]. In this context, the role of *data association* (i.e. the mechanism of association of sensory data to known memory items) in integrating the bottom–up and top–down information flows has been largely ignored. Data association algorithms from dynamic systems theory often involve mechanisms that predict future sensory stimuli in order to operate efficiently under real-world limited resources constraints [7]. At the same time, the biological brain is also thought to employ mechanisms such as prediction, anticipation and data association to survive in dynamic and partially known environments [23,40,56]. Despite this evident missing link, no concise models have been proposed so far, that use data association as a mechanism that links bottom–up and top–down information flows. Here we propose a modular cognitive framework that is based on probabilistic data association theory capable of integrating bottom–up and top–down information flows of attention, that is also plausible from a functional biological perspective. The proposed cognitive framework is tested on real-world artificial cognitive systems that act under limited resources constraints.

Building artificial sensori-motor systems has been the hardest benchmark for biologically inspired cognitive models and has played a key role in identifying the key mechanisms of cognition [62]. However, despite the fact that many cognitive architectures have been proposed over the past three decades, none of them account for the interplay between the bottom–up and top–down attentional mechanisms nor clarify the role of data association in this context [5,18,19,42,44,72]. Even though some cognitive architectures (e.g. the Distributed Adaptive Control (DAC)) have been tested in real-world applications and make concrete proposals for self-contained hierarchical cognitive systems, they do not incorporate an attention mechanism to enable the cognitive agent to act under limited resource constraints [72]. Some other architectures do propose top–down attentional mechanisms, but do not explicitly connect data association to attention and prediction mechanisms [18]. Yet, the integration of bottom–up sensory information and top–down knowledge using predictive mechanisms of data association remains unexplored.

Starting with the DAC architecture we propose an integrated cognitive framework to address this issue and we prove its feasibility by testing it in real-world and simulated environments. We propose *Prediction – Anticipation – Sensation – Attention – Response (PASAR) or DAC8*, a model of knowledge acquisition, representation and action selection for artificial autonomous systems. PASAR makes use of automated reasoning procedures involving anticipation of future stimuli and fusion of bottom–up and top–down information streams. Thus, data association serves as the underlying mechanism that integrates the key ingredients of cognition, namely attention, prediction, anticipation and action. The modular formulation of the PASAR architecture allows the dissection and evaluation of the contribution of its individual components. Thus, PASAR is a very valuable tool for a systematic evaluation of the interplay between memory, attention, prediction, perception and action. We demonstrate that PASAR can solve complex real-world and simulated robotic tasks. Our experimental results have strong implications on how anticipation, attention, sensation and action interplay. We show that a combined attentive, predictive and anticipatory system is clearly preferable to systems without those mechanisms when acting in dynamic environments under limited resources constraints. We also find that using a transient memory is beneficial to enhance the positive effects of attention, prediction and anticipation on system performance. Beyond providing a framework to investigate the interplay of the sub-components of cognition, we also demonstrate that PASAR serves as a useful tool to purposefully integrate perception, cognition and action in real-world behaving artificial systems.

2. Related work

Great efforts have been spent over the past two decades in discovering the brain mechanisms involved in integrating the bottom–up and top–down information flows in the brain in order to understand the interplay between prediction, anticipation, perception, memory and action [43]. Scrutinizing the highly hierarchical structure of the brain has been a starting point for various studies investigating these sub-components of cognition. Growing evidence from physiological research suggests that higher cortical areas of the brain are involved in a dialog with lower areas (mid-brain and superior colliculus) to solve low-level tasks, like the integration of multisensory information [66]. Recent neuroscientific and behavioral evidence also suggests that higher cortical areas of the brain *tamper* with the incoming sensory data to fit expectations by actively searching for relevant information [56]. This was also demonstrated earlier using evoked magneto-encephalography studies in humans that demonstrated that brain responses code for expected future stimuli [46,55]. Some researchers argue that higher cortical areas are involved in predicting future stimuli based on past experience and anticipate sensory events in time and space when solving complex sensori-motor tasks [23]. In this context, bottom–up and top–down mechanisms are highly relevant and have long been demonstrated to be at work in biological cognitive systems [14,33,58]. In the domain of attention, influential bottom–up processing models [33] and more recently top–down models [58,68] have been proposed. Some other theories propose the emergence of attention from knowledge representation itself, suggesting the interplay between subsystems at different hierarchical levels of the brain [10,20,71]. The potential neural correlates of the sub-systems responsible for multi-sensory integration and attention generation have been the subject of extensive research in psychology and cognitive neuroscience [28,30,60]. The interdisciplinary field of cognition and robotics research has greatly profited from the above findings about the key hierarchical and vertical brain mechanisms involved in biological cognition [75].

Based on empirical (anatomical, physiological and behavioral) evidence supporting the notion of layered hierarchical control systems, many dominant theories have so far been proposed to model a multi-layered framework of the brain, that has strong top–down and bottom–up information flows (see [62] for a review). When many of those models have been designed with the aim of building novel artificial systems, some others do not consider real-world performance issues. The subsumption architecture was proposed as a distributed architecture for building autonomous systems capable of sustained behavior in real-world environments [11]. The symbolic SOAR architecture has been under development since the 70s and aims at explaining reasoning from a psychological perspective [42]. ACT-R is yet another family of cognitive architectures, concerned primarily with modeling human behavior [5]. Other approaches such as ICARUS use notions of hierarchical relationship between objects to achieve problem-solving behavior [44]. Others have proposed biologically based hierarchical cognitive architectures emphasizing the notion of embodied cognition and linguistic interaction [19] or action imitation based on hierarchical representations of perception and action [18]. Also in this context, the Distributed Adaptive Control (DAC) framework has been proposed to accommodate perceptual and behavioral learning in artificial systems in a single framework [72] (see [43] for a review). Nevertheless, to our knowledge none of the above (and other) cognitive models address the issue of integrating bottom–up sensory information and top–down knowledge to deal with the limited resource constraints problem.

3. Research question

Although most of the above mentioned architectures have been highly influential in cognitive sciences and robotics research, none of them specifically addresses the limited resources constraints issue at perceptual, cognitive and motor action levels. Moreover, the role of data association as the integrative mechanism of bottom–up and top–down information flows of attention using predictive mechanisms has not been investigated. We also believe that such an integrative framework is necessary to understand the interplay between the different sub components of cognition and the contribution of each sub component to the whole artificial system in realistic tasks such as the ones performed by biological systems. The DAC architecture was proposed earlier for structuring perceptual and behavioral learning in three layers of control: reactive, adaptive and contextual [70,72] (see Fig. 1). Also recent versions of the DAC architecture address the issue of using predictions for sequence learning in robotic applications [20,71]. Here we further enhance the DAC architecture and propose PASAR, a concise and modular framework for integrating prediction, anticipation, sensation, attention and response. In

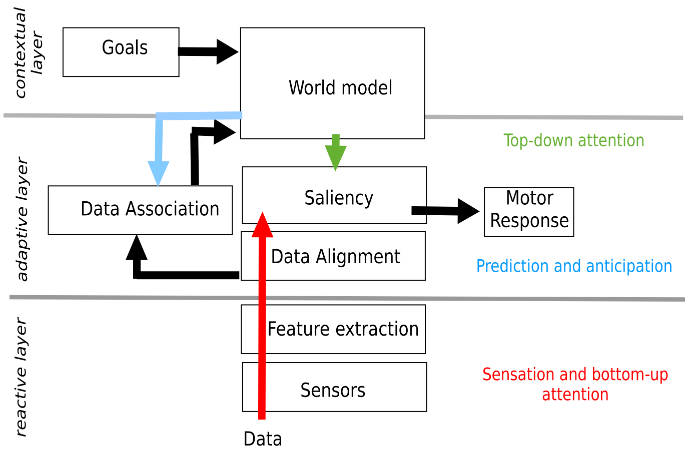

Fig. 1. **PASAR** is based on a three-layered distributed architecture: reactive, adaptive and contextual [72]. The reactive layer contains the physical *sensors* and the *feature extraction* mechanisms. The adaptive layer contains the *data alignment*, the *data association* and *saliency* computation mechanisms. The contextual layer contains the *world-model* and the *goals* of the system. The arrows indicate information flow, and the colored arrows indicate sensation and bottom–up attention (red), prediction and anticipation (blue) and top–down attention (green). The *motor response* is a result of the integration of the bottom–up and top–down saliency maps. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

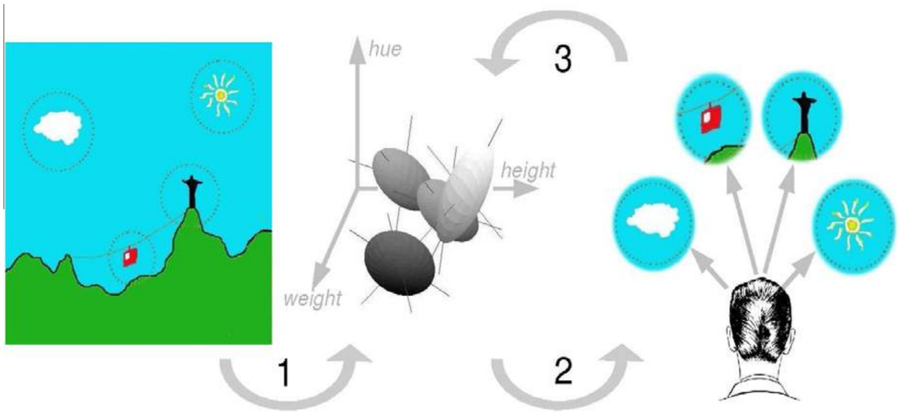

Fig. 2. Schematic of *world-model* and selective attention generation for a dynamic scenario.**Left panel:** A dynamic scene as perceived by an autonomous system. Four encircled objects (referred to as *concepts*) are perceived as closed entities by the artificial cognitive system. **Middle Panel:** Four *concepts* in $n = 3$ feature space with *hue*, *weight* and *height* as example features. The ellipsoids represent the covariance of the *concepts*. **Right Panel:** The top–down attention mechanism triggers an action that might have an immediate effect on the world model (arrow 3) as the sensory input is changed by the performed action.

our experiments in simulations and complex real-world robotic tasks, we address the question of how each sub-component of PASAR contributes to the overall performance in the given task and what insights we can gain about robotic control and cognitive mechanisms in general. More concretely, we want to understand if a complete sensori-motor system with attentional, anticipatory and predictive mechanisms performs better than incomplete subsystems in complex tasks. Also we address the use of a memory decay mechanism (forgetting) and the use of higher level knowledge for modulating bottom–up sensory information, as observed in biological systems [67]. Furthermore, we ask how the interplay of the different cognitive subsystems involved in solving complex tasks influences the task at hand. We also aim at providing a new hypothesis about the functional organization of biological cognition and also invitations to perform specific physiological and psychophysical experiments to support our findings in biological systems.

4. Research design

The proposed PASAR framework is embedded in the DAC architecture and encompasses three layers of control: reactive, adaptive and contextual (see Fig. 1). This layered hierarchy of control has been shown earlier to be optimal for solving diverse robotic sequence learning tasks [72]. The arrows in Fig. 1 indicate information flows. Sensory perception and the bottom–up saliency computation are indicated by the red arrow and are contained in the reactive layer. The adaptive layer contains the data alignment and the data association mechanisms. Also, the bottom–up and top–down attentional saliencies are integrated in the adaptive layer. The contextual layer accomplishes high-level knowledge acquisition in the world-model, the top–down attentional modulation of action generation (green arrow) and the anticipatory bias in data association (blue arrow). The individual components are elaborated in the following subsections.

4.1. Sensory data alignment

As seen in Fig. 1, the adaptive layer contains the data alignment mechanism. Data alignment is necessary to map different sensory coordinate systems to a common reference frame. Such aligned maps of multimodal sensors have long been discovered in the superior colliculus of the brain [67]. The superior colliculus, which has top–down connections from higher cortical areas, is considered as the primary domicile of sensory data association and appropriate motor action generation in animals [67]. It has been shown that the superior colliculus contains a *sensory map* for each sensor, on which the whole sensory space is represented on [35]. Whenever one of the sensory modalities is perturbed, biological systems have been shown to be capable of (re) learning such sensory representations using the representation of a non-perturbed sensory modality as the reference [39,74]. Analogously the DAC adaptive layer uses top–down information to learn sensory map alignments as shown in Fig. 1. PASAR is transparent to the specific learning mechanism itself, and we have demonstrated the capability of PASAR using two different learning paradigms, asymmetric Hebbian learning and Levenberg–Marquardt.[1]

4.2. World model

The artificial sensori-motor system acts in a complex world using multimodal sensors for perception where all aligned sensory information is used as input to the *world-model* (see Fig. 2). The *world-model* of the contextual layer contains all

[1] See Supplementary material, section B for implementation details.

the high-level knowledge about the external world, acting as a memory mechanism (Fig. 1). This memory is organized in terms of individual memory items, further referred to as *concepts*. *Concepts* are basically individual memory items and are formalized as vectors in a high dimensional space. The *world-model* is seen as the result of a memory mechanism for acquiring and maintaining knowledge about different *concepts* either provided by bottom–up multimodal sensory signals or by long-term-memory retrieval. The uncertainties about the concepts are defined by their covariances. The covariance of the *concepts* represent uncertainties about the spatio-temporal (in) congruence of the incoming sensory data and their predictions. Also, as resources are limited, only a fixed number of concepts can be maintained in a finite memory. For this reason we employ a *forgetting* mechanism of concepts. *Concepts* are *transient* as they are forgotten with the passage of time, where we define *forgetting* as a linear growth function of the covariance of the *concepts*.

4.3. Prediction, anticipation and sensation

Given the limited resource constraint of self-contained cognitive systems that act in a dynamic world, predictions about future sensory stimuli are necessary to be able to sense and act in real-time [37]. Given the above *world-model* with *concepts* and their covariances, we discuss here how predictions of future stimuli and anticipations are computed. *Prediction* refers to the mechanism of predicting future stimuli based on recently perceived stimuli related to a *concept*. Harnessing such predictions, sensory stimuli *associated* to *concepts* can be *anticipated* in a spatio-temporally constrained region of sensory space. In this sense, multimodal sensation involves prediction, subsequent anticipation and finally data association.

Bayesian models have often been proposed in this context for neural coding of sensory input [38], attention [32], sensorimotor learning [41] etc. Here we use the so-called Joint Probabilistic Data Association (JPDA) algorithm, which uses Bayesian inference to compute the most probable sensory data to memory item associations [7]. JPDA is a very interesting candidate for data association in our context as it intrinsically sets a boundary for data association, which is very useful for systems with limited sensory, cognitive and motor resources. JPDA is a powerful tool for solving data association problems, which arise in many applications such as computer vision, surveillance, mobile robotics etc. The JPDA algorithm uses the notion of validation gate to restrict the association of data to concepts. JPDA is applicable to non-linear filters and our results can be easily generalized [7]. In the following, each *concept* is a well defined multidimensional vector of sensory data in a high dimensional space, that depends on the type of the multimodal sensory input of the system. Let $\mathcal{K}$ be the number of *concepts* in memory. The state dynamics of *concept k* is modeled as

$$x_{t+1}^k = A_t^k x_t^k + \psi_t^p \tag{1}$$

for $t = 1, 2, \ldots$ where $x_t^k \in \mathbb{R}^{n_x}$ is the state of *concept k* at time t.ψ_t^p is process noise with covariance matrix Q. Let $y_t^j \in \mathbb{R}^{n_y}$ be the jth observation at time t for $j = 1, \ldots, n_t$. The measurement model is

$$y_t^j = Hx_t^j + \phi_m \tag{2}$$

where ϕ_m is a white Gaussian measurement error with covariance R_k. Given this Gaussian noise model and assuming linear dynamic and measurement models, we use a Kalman filter for state approximation of *concepts* [36]. In this sense, the Kalman filter is an iterative mechanism that allows us to make predictions of future sensory input as well as to use the sensory input to correct the state of the world-model. Nevertheless, it should be noted that Kalman filter is not proposed here as a model of the brain, but serves solely as one possible implementation of the involved prediction mechanism.

Using apriori knowledge about the world (e.g. state transition matrix, process noise and measurement matrix in the case of the Kalman filter) and the current state of the world-model, a prediction is made for each *concept*. At time step t, for each *concept k*, we compute the state prediction, its covariance and the measurement prediction as follows

$$\tilde{x}_t^k = A\hat{x}_{t-1}^k \tag{3}$$

$$\widetilde{P}_t^k = A\widehat{P}_{t-1}^k A^T + Q_{t-1}^k \tag{4}$$

$$\tilde{y}_t^k = H\tilde{x}_t^k \tag{5}$$

Anticipation refers to harnessing the prediction to expect stimuli in a spatio-temporally constrained region of the feature space for the known *concepts*. Given the computed measurement prediction for *concept k*, we now compute an anticipatory field (validation gate) for the same *concept*. For each observation j the Kalman innovation (i.e. the prediction error computed as the difference between the prediction and the actual stimulus) and its covariance (for details see [36]) with respect to a *concept k* are:

$$v_{jk} = y_t^j - \hat{y}_t^k \tag{6}$$

$$S_{jk} = H\widetilde{P}_t^k H^T + R_k \tag{7}$$

An anticipatory field, or validation gate, is defined for each *concept* as follows:

$$v_{jk}^T S_{jk}^{-1} v_{jk} < \epsilon \tag{8}$$

Since the weighted norm of the innovation that defines the validation gate is chi-square distributed with the number of degrees of freedom equal to the dimensionality of the measurement, the threshold ϵ is obtained from the tables of a chi-square distribution [7]. The size of the obtained anticipatory field can be modulated by higher-level mechanisms by varying this threshold. A validation gate for each *concept* dimension is computed using the Kalman innovation of new observations as in [7].

Sensation or perception is traditionally defined as the raw sensory input available to the system at a given time [58]. Nevertheless, here we use the operational definition of *sensation* as the state estimation of *concept k*, using the given sensory input, prediction and anticipation. Given multiple *concepts* and multiple stimuli, the problem of data association (see Fig. 1) naturally arises [7]. For each *concept*, only observations inside its validation gate are associated to it. The JPDA algorithm enumerates all possible associations between observations and *concepts* at each time step. The association probability β_{jk} stands for the probability that the jth observation originated from the kth *concept*. The *concept* state is estimated by the Kalman filter and this conditional expectation of the state is weighted by the association probability. x_t^k indicates the state of *concept k* at time step t, whereas ω_{jk} represents the association event where the observation j is associated to *concept k* and $Y_{1:t}$ stands for all the observations from time step 1 to time step t. ($Y_{1:t}$ stands for all observations from time step 1 until time step t, and is constrained by the available limited memory. We discuss below how this bottleneck is handled using approximation strategies in variants of the JPDA algorithm). The state of the *concept* can be estimated as

$$E\left(x_t^k|Y_{1:t}\right) = \sum_{\omega} E\left(x_t^k|\omega, Y_{1:t}\right)P(\omega|Y_{1:t}) \tag{9}$$

$$= \sum_{j} E\left(x_t^k|\omega_{jk}, Y_{1:t}\right)P(\omega_{jk}|Y_{1:t}) \tag{10}$$

where ω_{jk} denotes the association event when the observation j is associated to *concept k* and ω_{0k} denotes the event that no observation is associated to *concept k*. Therefore the event association probability is

$$\beta_{jk} = P(\omega_{jk}|Y_{1:t}) \tag{11}$$

The computation of β_{jk} requires a summation over the posteriors and its exact calculation is NP-hard [16]. We implemented a Markov Chain Monte Carlo (MCMC) method to compute β_{jk} in polynomial time similar to the proposal in [57].[2]

Time dependent forgetting is achieved by linearly growing the covariance of the represented concepts with the passage of time. This allows to remove the representation of concepts in memory if they are not updated by associations to novel stimuli, similar to the rehearsal-decay paradigm for memory [6]. In short, the *world-model* acts as a memory mechanism containing *concepts*. The *prediction*, *anticipation* and *sensation* mechanisms act on *concepts* in the *world-model*. Next we discuss the integration of top–down and bottom–up *attention* into the same framework.

4.4. Top–down and bottom–up attention

Having elaborated on the generation of the *world-model* containing the *concepts*, the data association, prediction and anticipation mechanisms, we discuss how attentional mechanisms can be integrated in the same framework. In our context, attention is basically the process of focusing the sensory resources to a point in space. We intend to phrase attention in terms of bottom–up and top–down processes as employed by the brain [33]. The focus of attention can simply be non-specific spacial positions in case of bottom–up attention, but also the specific spatial location of a real-world object which has a certain memory item (see Fig. 2 right panel). Recent psychophysiological research suggests that selective attention is load-dependent [60]. Artificial systems with limited and/or shared resources could benefit from such a load-dependent protocol. Our architecture makes use of a load-dependent push–pull mechanism, by means of a competition between distinct feature saliencies of *concepts*, the winner of which decides which actions or sensory events the system decides to dedicate its resources to. In PASAR, low level feature filters extract features of incoming multimodal stimuli. For vision, a number of feature filters like color, orientation, luminance etc. are implemented. The combination of these filter outputs give rise to a bottom–up saliency map based on the Itti and Koch model [33]. At the same time, top–down attention is triggered by similarity interference, i.e. when similarity between memory *concepts* is the major memory impairing factor of identification, similar to some earlier proposals [54]. Similarity interference between *concepts* is measured using the Mahanalobis distance between the *concepts* [49]. Top–down and bottom–up saliency maps are combined and a winner-take-all (WTA) neural network computes the most salient attentional *spotlight*, which is then used to trigger an action (Fig. 3). See also [51] for more details.

4.5. Motor response

Until now, we have discussed the mechanisms of *world-model*, *prediction*, *anticipation*, *sensation* and *attention* in the PASAR framework. However, to act in a dynamic world, the sensori-motor system needs to perform motor responses (Fig. 1). To choose the optimal motor response we consider the general problem of an efficient energy consumption strategy [8]. In this formulation, the artificial autonomous system has to optimize the *utility* of its actions while at the same time minimize its limited energy consumption. As the utilities of future actions are apriori not known, predictions of the same are necessary. Here we formulate this in the context of a Bayesian framework based on transient memory, prediction and attention.

[2] See details of the MCMC implementation in Supplementary material, section A.

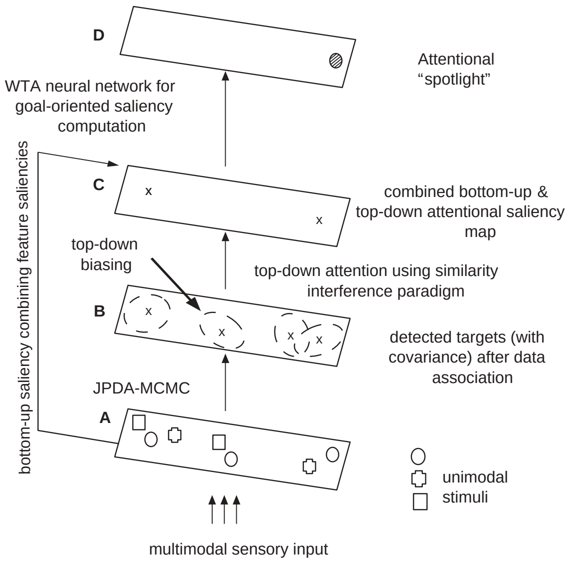

Fig. 3. Saliency computation from multimodal sensory input: The multimodal sensory input (**A**) is associated to existing targets by means of the anticipatory fields of JPDA (**B**). The top–down biasing of the anticipatory fields using the world-model is applied before data association. The bottom–up and top–down saliency maps are combined using a common neural group onto which both the above saliencies are mapped (**C**). A winner-take-all WTA neural network computes a single winner from this map (**D**) and an appropriate action is triggered.

In this framework, the *attentional saliency* of a *concept* is defined as the perceived relevance of the stimuli associated to this *concept*. Note that although this is a simplification that reduces all forms of attention into a single variable, it allows for testing the contribution of attention to overall performance in a well-defined testbed. We assume that the *utility* of an action is predefined for a given task. E.g. in our multi-robot testbed the utility of a *concept* is defined as inversely proportional to the perceived charge of the robot (i.e. high utility for a low charge robot and vice versa). PASAR computes the predicted utility distribution of actions from the individual utilities of single *concepts*. We write Θ_s^t for utility of a certain *concept s* at time *t*, and for computing the utility distribution we consider the following conditional probability distribution defined as the predicted utility:

$$\text{Predicted utility of concept} = P\left(\Theta_s^t | F_s^t\left(\Theta_s^{t-1}\right) A_t(s)\right) \tag{12}$$

where $F_s^t\left(\Theta_s^{t-1}\right)$ and $A_t(s)$ are two time-dependent functions that weight the *concept s*. For example, $F_s^t\left(\Theta_s^{t-1}\right)$ evaluates the temporal weight of the *concept*: if there is at least one stimulus associated to this *concept* currently, it has the highest temporal weight and decays linearly otherwise with time. Such a mechanism allows for a simple time-dependent forgetting. $A_t(s)$ evaluates the attentional saliency of this *concept* proportional to the perceived utility of the *concept*. By computing the joint distribution of the predicted utility probabilities for all *concepts*, the system can perform the action with the highest predicted utility. We elaborate the update mechanism of this predicted utility distribution. Lets assume that we can compute predicted utility probabilities of individual *concepts* as shown in Eq. (12). Given these individual *concept* utilities, we are interested in the total predicted utility distribution:

$$\text{Predicted total utility} = P(\Theta^t | F^t(\Theta^{t-1}) A_t) \tag{13}$$

We express this probability as the normalized sum of probabilities of individual utilities:

$$P(\Theta^t | F^t(\Theta^{t-1}) A_t) = \sum_s P(s) P\left(\Theta_s^t | F_s^t\left(\Theta_s^{t-1}\right) A_t(s) S\right) \tag{14}$$

where S is a random variable so that $S \in 1 \ldots n$, n being the number of *concepts* and $P(s)$ indicates the probability of the utility of this *concept*. Normalization is straightforward as $P(s)$ is uniformly distributed over all *concepts*. PASAR then either selects the action of maximum utility or draws from this distribution, where the former can be seen as *exploitation* of the world-model and the latter as *exploration*. We are therefore interested in the following probability distribution of actions:

$$P\left(Action | F_s^t\left(\Theta_s^{t-1}\right) A^t(s)\right) \tag{15}$$

This probability distribution can be computed using Bayesian inference, given apriori information about the environment PASAR is acting in.

The formalizations in this section allow us to test the contributions of the individual components of PASAR, prediction, attention and memory decay, to the performance of the system. In summary, we have elaborated on how an optimal motor

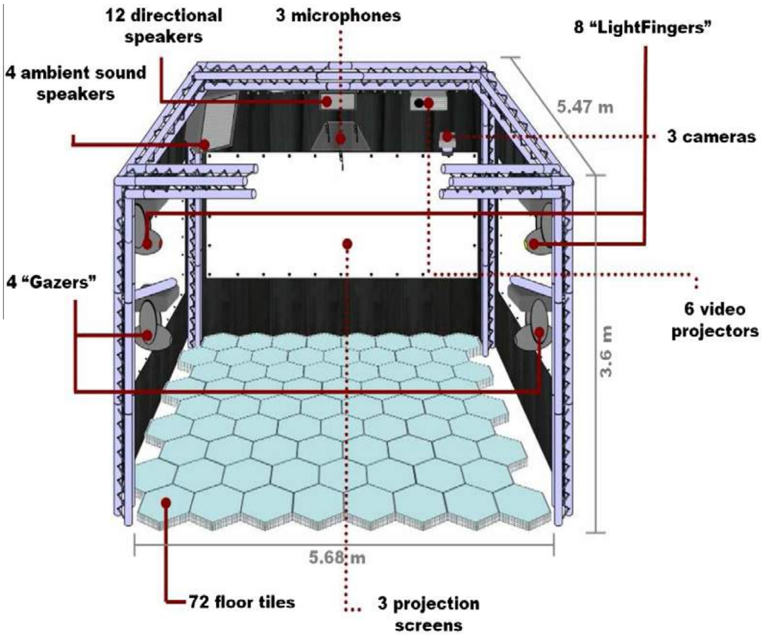

Fig. 4. The eXperience Induction Machine (XIM) can be considered as an artificial organism in the shape of an environment, that has its own goals and expresses its own internal states. It comprises a pressure sensitive floor, overhead and pan-tilt cameras (gazers), movable lights (light-fingers), triplets of microphones for sound recognition and localization, projection screens and ambient and directional sound output (adapted from [50]).

response is chosen by the system using a prediction, anticipation, sensation and attention. In the next section, we define our experimental setup, with which we want to quantify the contribution of each of the above sub-components to the overall system performance.

5. Experimental setup

In the above sections we have elaborated on the PASAR model and how the different sub-components (prediction, anticipation, sensation, attention and response) are formalized. Here we discuss the rescue robot simulation and the real-world robotic testbeds, which we use to evaluate PASAR's performance and the interplay of the sub-components.

5.1. eXperience Induction Machine (XIM)

XIM is a physical space, which is part of the Persistent Virtual Community (PVC[3]) where groups of real, remote and synthetic characters interact with each other. XIM comprises a pressure sensitive floor, overhead, infrared and controllable pan-tilt color cameras, moving lights, triplets of microphones for sound recognition and localization, projection screens, and also ambient and spatialized sonification. On the projection screens, the virtual world of PVC is made visible to the real visitors of XIM. XIM is about 25 square meters and allows several humans to be active in it simultaneously (Fig. 4). XIM, as an artificial autonomous entity, is a suitable testbed for PASAR as it comprises of multimodal sensors and effectors in a scalable and controlled environment for solving complex collective mixed reality interaction tasks (Fig. 5). The accurate tracking of real objects in the XIM physical space is a requirement for meaningful interaction scenarios among the different types of users. PASAR is tested in XIM for solving this task.

We use the IQR system for real-time neuronal simulations of the winner-take-all network implementing the attention system for driving the movable pan-tilt cameras in XIM [9]. For image processing we used the OpenCV library [1], and for the human torso detection in the mixed reality space testbed we used the OpenCV Haar classifier method [73]. PASAR implementation for the mixed reality testbed runs 7 applications developed in C++ on 3 Intel (R) Core (TM)2 Duo CPU 2.66 GHz machines with the GNU/Linux Suse10.3 operating system, which communicate with each other using UDP sockets.

5.2. Rescue robot simulation

In this testbed we consider the following multi-robot scenario: *N* number of robots are on a common mission in a given environment. The individual robots move around in a given common field solving a given task (e.g. de-mining). One of the

[3] http://specs.upf.edu/projects/337.

Fig. 5. XIM Multiuser Interaction Scenario: Multiple real users interacting with each other in a mixed reality Pong game, which is one of the scenarios used to evaluate PASAR's performance. Also remote users take part in the same interaction by logging in from remote machines into the virtual world and they are represented by avatars on the projection screens (adapted from [31]).

robots, named PASAR, has the specific task of rescuing expired (out of charge) robots. To achieve this, PASAR first has to localize the expired robots using its sensors, approach them and finally recharge them so that they can continue on their missions. PASAR is equipped with a limited number of distance measurement sensors like sonar and laser range scanners, which are used to scan the environment and localize the robots to recharge. We simplify all these sensors in the form of a circular *perceptive field* around PASAR, inside which it perceives its co-robots. From time to time, PASAR has to return to the base station to recharge itself. We implemented this testbed in a simulation with $N = 10$ robots (Fig. 6). *Concepts* in this testbed are the robots involved in the common mission. For the Bayesian inference computations we used the ProBT Bayesian library [2]. The simulation runs on a single Intel (R) Core (TM)2 Duo CPU 2.66 GHz machine with the GNU/Linux Suse10.3 operating system.

6. Data collection

Given the PASAR model and the testbeds, here we discuss the data we collect in the two testbeds.

6.1. XIM

In this testbed, learning sensory map alignment, creation and maintenance of a world-model under limited resources constraints when acting in an unknown environment and attentional mechanisms for active information acquisition are tested. This testbed deals with real-time issues of sensory map alignment learning, creation and maintenance of a world-model and attention.

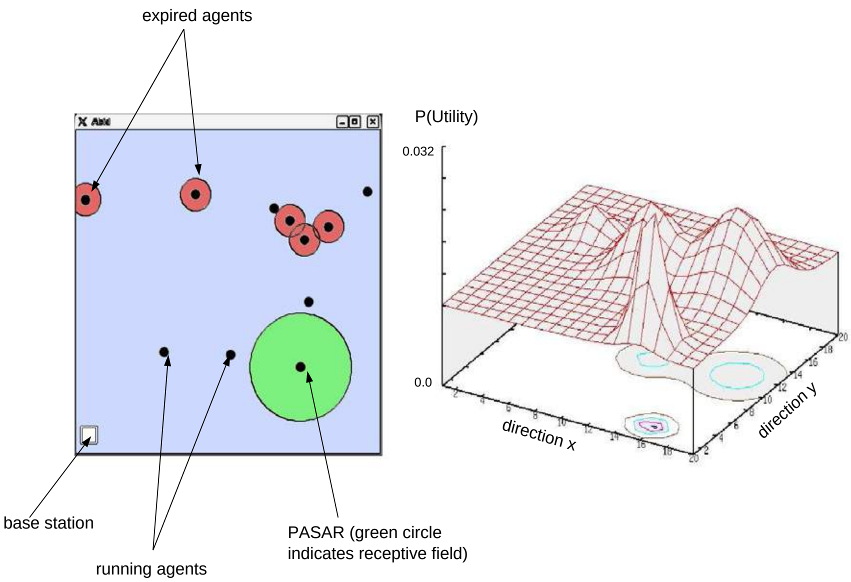

Fig. 6. PASAR multi-robot rescue scenario, left Snapshot from PASAR Experiment: The bigger circle surrounding PASAR shows sensory range. Everything outside this range is not perceived by PASAR. The smaller circles surrounding some of the agents indicate that the corresponding agents are expired. The base station, where PASAR recharges itself, is indicated by the rectangle on the left bottom. Right example of the predicted total utility as a probability distribution. PASAR either goes to the location with the highest probability (exploitation phase) or draws form this distribution (exploration phase).

The sensory map alignment of the overhead camera and floor modalities are learned online by a person walking in XIM.[4] For evaluating the construction and maintenance of a world-model we look at the problem of multi-person tracking. Tactile floor tiles and overhead infrared cameras are the sensors used for this. A *concept* here is a single human in XIM. We use a 8-dimensional space for the *concepts* containing the x-y positions, x-y velocities, x-y accelerations, hue of the outfit and weight. To test the precision of the world-model during interactions involving different difficulty levels of tracking, a number of experiments are performed.[5] Bottom-up and top–down attention for action generation is tested using the controllable pan-tilt cameras for collecting hue information about tracked persons. Using this active recruitment mechanism we analyze ID maintenance in four different interaction scenarios.[6] To evaluate the scalability of the system, we also implemented a simulation of the XIM environment to test PASAR in the above tasks with a much higher number of persons than was possible in the real setup.

6.2. Rescue robot simulation

PASAR is tested in a multi-robot simulation to assess the optimal usage of the acquired world-model in multiple goal decision making. This testbed, in contrast to the previous one, allows for testing how to capitalize on the availability of such a world-model for generating optimal actions in a partially observable world. In particular, we will use this second testbed for evaluating the contributions of the individual components of PASAR to the overall performance.

In this context, the *attentional saliency* of a *concept* is defined as the perceived utility of the stimuli associated to this *concept*. Therefore, the *attentional saliency* is defined to be proportional to the detected charge of a *concept*, giving a high utility for re-approaching nearly expired *concepts*. PASAR computes the predicted utility distribution of actions from the individual utilities of single *concepts*. Performance is measured by the total expiry time in seconds of the robots. I.e. the lower this value, the better the performance. For each test case 100 trials were performed. Each trial lasted 160 s. The expiry time of the normal robots was 30 s and that of the PASAR rescue robot was 60 s. Solving this multiple goal task involves goal-driven selective attention generation for attending to the most relevant sub task at the moment and maintaining a dynamic transient world model, which is used to compute the optimal action in the Bayesian sense. We formulated an optimal Bayesian decision making method for generating actions based on the world-model. PASAR computes the predicted utility distribution of going to a certain point (x,y) from the individual utilities of the *concepts*.[7]

In order to evaluate the contributions of prediction, attention and memory decay, we perform the following experiments. First we evaluate the performance of the whole system using the complete PASAR framework for the robot rescue task (*COMPLETE*). Note that for this case we use the optimal memory decay rate, as discussed in Section 5.2, and an adaptive exploratory behavior of PASAR. The optimal memory decay rate is set to be the memory decay for which the highest performance was observed for the *COMPLETE* test case (value 13, see Fig. 14). With adaptive exploratory behavior, we mean that PASAR drew from the predicted utility distribution much more often at the beginning of the run, than towards the end. This allowed PASAR to explore the world at the beginning, when the world-model did not contain any *concepts*, and to exploit this acquired knowledge more towards the end of the run. This behavior is achieved simply by setting a draw probability that correlates negatively with time in a linear fashion. Next we evaluate the performance of the system without the use of attention, i.e. we set the attentional saliency of all perceived *concepts* to be the same (*NO-ATT*). Further we evaluate the performance of the system without memory decay and without attention (*NO-DECAY-NO-ATT*). We also test the case without memory decay, but with attention (*NO-DECAY-ATT*). Finally we evaluate the performance of the system also when not using the utility predictions of the world-model. I.e. the rescue robot randomly wanders around the arena to find and recharge expired robots (*RAND*).

7. Data analysis

For all offline analysis we used Matlab (R) (2007a, The MathWorks). For multiple error comparisons we use Tukey-Cramer multiple comparison method $p < 0.05$. The hue bin comparisons were done using the student's t-test $p < 0.05$. For the Bayesian inference computations and analysis we used the ProBT Bayesian library [2].

8. Results

8.1. Mixed reality space XIM testbed

In the XIM mixed reality space testbed, we test the online learning of sensory maps, use of high level knowledge for world-model maintenance, the use of attentional resources for triggering motor commands for active information acquisition and the use of the same for correction of the world-model.

First, we consider sensory map learning using the PASAR adaptive layer. The tactile floor delivers low resolution but precise position data and therefore is used as the reference for aligning the higher resolution sensory map of the overhead

[4] See Supplementary material, section B., for details on sensory map alignment learning.
[5] See Supplementary material, section D, for details on multi-person tracking experiment in XIM.
[6] See Supplementary material, section E, for details on attention trigger in XIM.
[7] For the elaboration of how the general PASAR Eqs. (12)–(14) are applied to the multi-robot testbed, see Supplementary material, section C.

infrared tracking camera placed arbitrarily in XIM. The procedure for this learning consists of a person walking freely in XIM while the tactile floor and the overhead infrared camera tracking data are used by the PASAR adaptive layer. See Section 6.1 for details of data collection. This procedure drastically reduces the camera perspective and distortion errors (from 75 to 26 cm) in the overhead camera tracking (Fig. 7 A,B, Tukey-Cramer multiple comparison, $p < 0.05$). After this initial learning, the overhead camera tracking still has relatively high error along the periphery of the tracked area (mean error 26 cm, Fig. 7B). The PASAR contextual layer fuses the multimodal data, considering prior knowledge such as the increased reliability of the tactile floor data along the periphery of the space. Such knowledge is used to modulate the validation gate of the floor modality accordingly (the larger the validation gate, the more reliable the modality). This gives less than 15 cm tracking position error (Tukey-Cramer multiple comparison, $p < 0.05$) (Fig. 7C). This demonstrates the use of top–down attentional bias using the validation gate thresholds (see Eq. (8)).

The PASAR world-model contains the high dimensional information about the tracked humans. To test the precision of the world-model during interactions involving different difficulty levels of tracking, a number of experiments are performed. ID maintenance is analyzed for four different interaction scenarios: *exploration, energy, center of mass, pong game*. The results in the four different scenarios show high precision of the world-model even for challenging tracking conditions involving occlusions, clutter and high movement speed (about 88% correct ID resolutions in average for 5 persons) (Fig. 8). To evaluate this with higher number of tracked objects, we use a simulation of the XIM persons and use PASAR for tracking. The performance falls with the number of tracked persons as expected for up to 20 persons. For each number of persons, we perform 20 trials. The computed mean and standard deviation are depicted in Fig. 9.

The idea of using apriori information is to be able to modulate the association probabilities of sensory data to tracked persons using apriori information about the world. We use the apriori knowledge that spatio-temporal proximity of multimodal sensory data has to be weighted more. Here the camera data that is proximal to floor data in space and time is weighted more, i.e. the association probabilities of such data to *concepts* are higher. Using this method improves the ID resolution in all interaction scenarios and this strongly supports the usefulness of top–down modulation of sensory data for enhancing data association performance (Fig. 10).

The performance falls with the number of simulated persons as expected (Fig. 11). Nevertheless, the performance is significantly higher than when not using the apriori knowledge for up to 12 persons (Tukey-Cramer, $p < 0.05$).

PASAR's just-in-time attention trigger allows to drive the pan-tilt cameras, which is a limited resource, to follow moving persons in XIM and extract hue information at the right moment (Fig. 12). The top–down attention for such deployment of movable cameras and moving lights is triggered as discussed in the Section 6.1, using the similarity interference paradigm. Sample images from the four moving cameras for two subjects in XIM are depicted in Fig. 12. We evaluated the similarity of the computed mean hue for two different subjects. The computed hue bin means were significantly different (student's t-test, $p < 0.05$) (Fig. 13).

Although hue extraction is a simple feature extraction method, it serves as a proof of concept for motor action generation from attention. By actively collecting hue information of subjects in the space, their IDs can be corrected. Latency for the recuperation of IDs after an ID confusion, using this attentional deployment of the controllable pan-tilts are shown in Fig. 13 for two subjects acting in the space simultaneously for 10 min. Mean time for ID recuperations is 17 s after confusion.

8.2. Robot Rescue Simulation Testbed

In this second testbed the objective is to evaluate the exploitation of a world-model for the generation of actions maximizing a given utility. Also we evaluate the contributions of the different components of PASAR, prediction, attention and memory decay, to performance. See Section 6.2 for details of the experiment. For each test case mentioned in Section 6.2

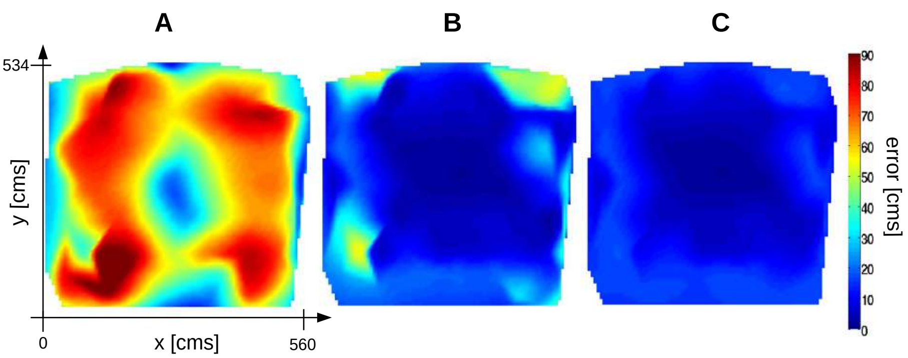

Fig. 7. Sensory Map Learning Using PASAR Adaptive Layer: A shows the error map in [cm] for the overhead infrared camera tracking before learning (mean 75 cm). High errors are due to camera perspective and distortion errors. **B** shows the error after online learning with a single user in XIM (mean 26 cm). **C** The validation gates of the *concepts* (i.e. the tracked persons) for the tactile floor modality are made larger along the periphery of the space. This is helpful as the floor data along the periphery is less cluttered and more reliable than in the center of the space. Using this top–down bias in PASAR improves the tracking error significantly (mean 14 cm). The error distributions decrease significantly from *A* to *B* to *C* (Tukey–Cramer multiple comparison, $p < 0.05$).

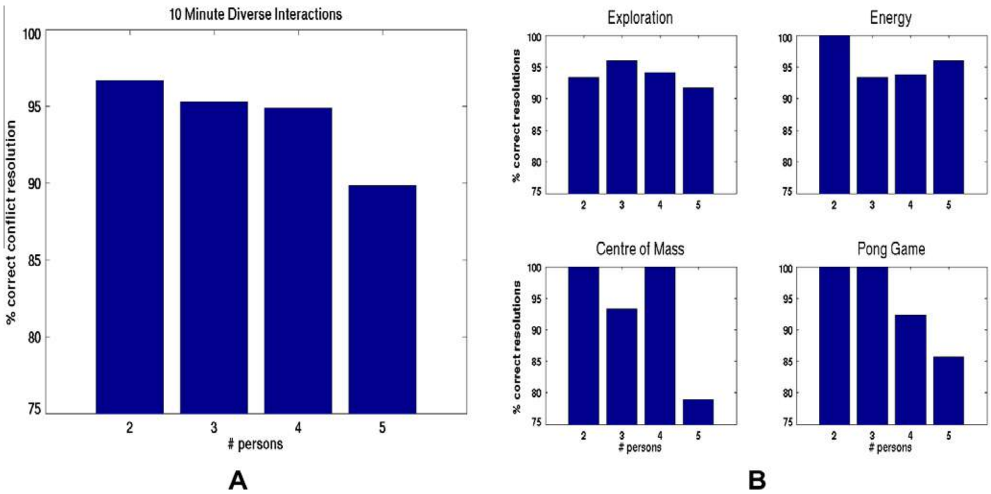

Fig. 8. PASAR for multimodal multitarget tracking: A shows correctness of ID resolution in cluttered situations of tracking with 2 to 5 persons freely interacting in XIM. **B** shows ID resolution accuracy for different interaction scenarios (see text for further details).

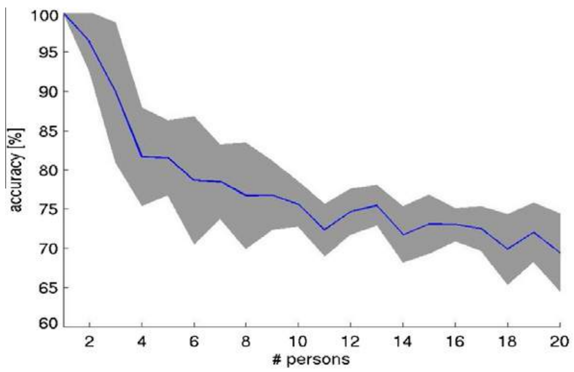

Fig. 9. PASAR for multimodal multitarget tracking: ID resolution performance percentage as a function of the number of objects tracked in XIM simulation.

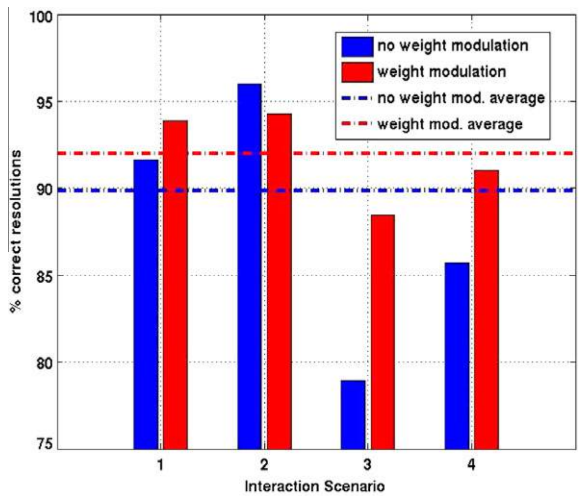

Fig. 10. Usage of apriori knowledge: spatiotemporal congruence of multimodal data in XIM, multimodal data that is proximal to data of another modality in space and time has an added weight. The scenario numbers 1, 2, 3, 4 refer to the interaction scenarios *exploration, energy, center of mass, pong game* respectively.

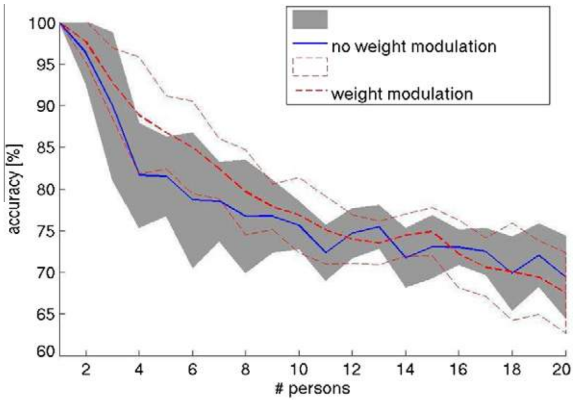

Fig. 11. Usage of apriori knowledge: spatiotemporal congruence of multimodal data in simulation: The accuracy of ID resolutions in percentage is depicted as a function of the number of tracked objects in a simulation of the XIM tracking scenario. The accuracy when using the apriori knowledge to weight modulate data and when not using it is shown.

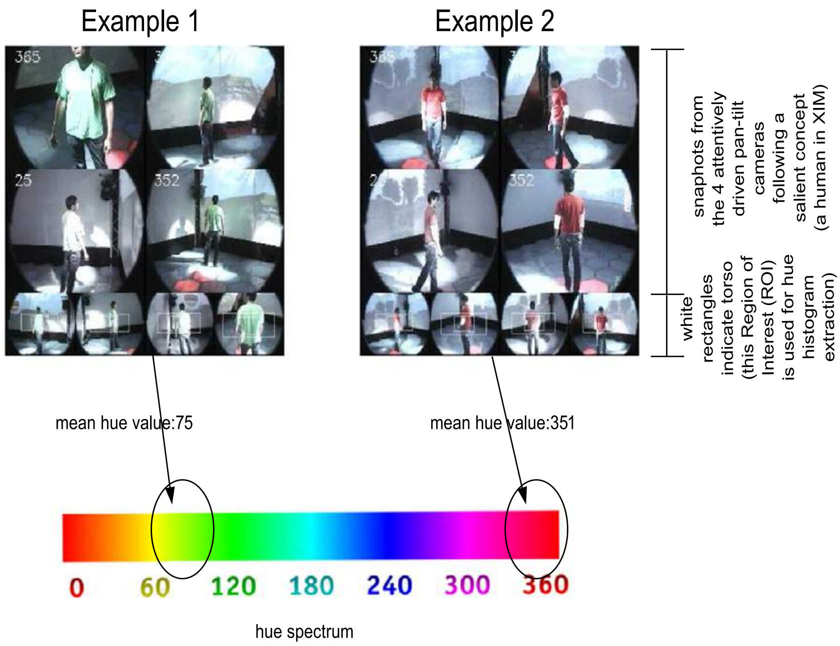

Fig. 12. Attention-guided feature extraction under noise and limited resources constraints: PASAR attention system follows subjects in XIM and hue extraction from torsos. Two such attempts are shown (green and red). The snapshots are from a moving camera image and the bottom indicates the images used for hue extraction. The Region-of-Interests (ROIs) of the images used to extract the hues are indicated by the white rectangular boxes in the smaller images at the bottom. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

trials were performed. Each trial lasted 160 s. The expiry time of the normal robots was 30 seconds and of the PASAR rescue robot was 60 s.

Fig. 14 shows the total expiry time of agents as a function of memory decay rate. It shows that neither a too low nor a too high memory decay rate is good for performance. In the *COMPLETE* case we achieve a mean performance of 100.6567 s and standard deviation 42.4679. In the *NO-ATT* case we achieve a mean performance of 131.0965 s and standard deviation 43.9256. In the *NO-DECAY-NO-ATT* case we achieve a mean performance of 249.5249 seconds and standard deviation 55.5471. In the test case without memory decay, but with attention (*NO-DECAY-ATT*) we achieve a mean performance of 248.0998 s and standard deviation 62.9844. Finally, in the random case *RAND* (i.e. without using predictions of utility) a mean performance of 145.9758 s and standard deviation 53.1304 is achieved.

We summarize the following:

(1) using the predictions of the world model together with attention and a good memory decay rate performs the best (Fig. 15, *COMPLETE*)

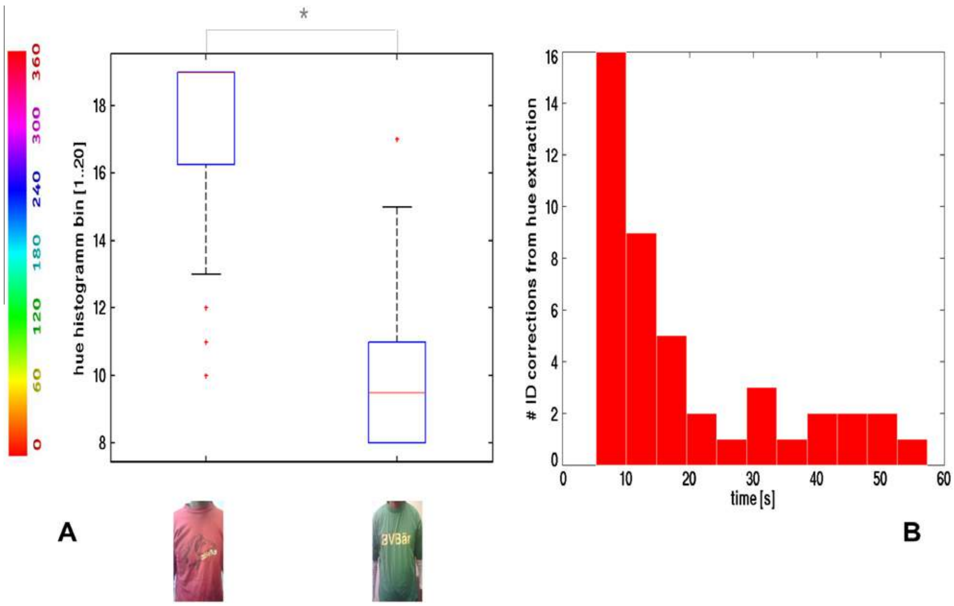

Fig. 13. Attention-guided feature extraction under noise and limited resources constraints: (A) Subjects with two different hues are identified as belonging to two different hue bins with a significant difference (student's t-test, $p < 0.05$). (B) Latency of recuperation: Latency of recuperation of IDs using hue feature extraction after an ID confusion (mean 17 s). The data shown is extracted from two subjects acting simultaneously in XIM for 5 minutes.

(2) Using attention is better than not using it (Fig. 15, *COMPLETE* vs. *NO-ATT*)
(3) Using memory decay has a significant effect on performance (*COMPLETE* and *NO-ATT* compared to *NO-DECAY-NO-ATT* and *NO-DECAY-ATT*).
(4) Further, attention does not seem to help performance if there is no memory decay (*NO-DECAY-ATT* vs. *COMPLETE* and *NO-ATT*).
(5) Using utility prediction does not help if no memory decay is used (*RAND* vs. *NO-DECAY-NO-ATT*).

9. Discussion

We primarily were interested in the mechanism generally known as *attention* and its bottom–up and top–down information pathways, that is employed by the biological brain to achieve selectivity at different levels of the perception-cognition-action cycle. Attention is thought to be a core mechanism that enables biological systems to excel in solving simple and complex tasks in dynamic environments facing limited resource constraints. We suggested data association as the mechanism to integrate bottom–up and top–down information flows of attention, and proposed an integrated and modular cognitive model.

Here we discuss the concrete implications of the present model and compare its merits and limitations with similar proposals. Highly influential general models of cognition have their own particular strengths, e.g. in powerful symbol

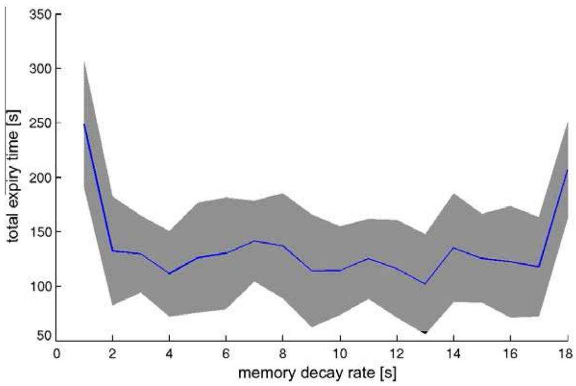

Fig. 14. Plot shows the total expiry time of robots as a function of the memory decay rate used in seconds. Decay rate refers to the time in seconds in which the *concept* variance falls to a predefined baseline.

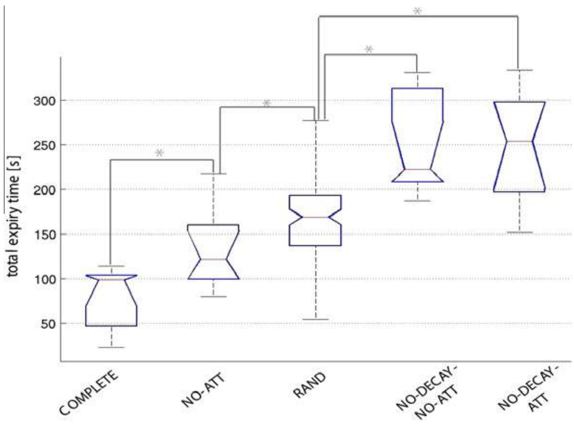

Fig. 15. Total expiry time for different test cases: Performance of PASAR in different test cases. Asterisks indicate significant difference of means (Tukey-Cramer multiple comparison, $p < 0.05$). *COMPLETE* test case performs the best.

manipulation and automation mechanisms [5], extremely versatile knowledge representation schemes [42], primacy of action and perception over cognition [44], language processing [19], perceptual and behavioral learning [70] etc. However, unlike other cognitive architectures, PASAR concretely provides a generic framework for integrating sub components of cognition and testing their interplay based on our current understanding of the brain. PASAR thereby makes use of a so-called *forward model* by harnessing predictions and anticipations [37]. In this context, prediction is essentially a problem of matching the state of the world as represented in the memory (the current situation of each *concept* in the *world-model* at a single point in time) to that produced by the model. Anticipation is essentially the process of expecting stimuli or events in the spatiotemporal vicinity of the predicted state. It is thought that something similar to such *state-matching* happens in the brain, through an ability known as *theory of mind*, which attributes mental states (beliefs, desires, intentions, etc.) to others and uses them to understand how others will behave [61]. Some brain theories previously have modeled conscious perception using the same principle of minimization of sensory prediction error [17,24]. PASAR integrates the different components of a cognitive system into a single framework embedded in a data association mechanism, providing a clear hypothesis about the intricate role of data association in a self-contained hierarchical cognitive system.

In the following we first discuss the implications of our results for some other recent proposals of integrated cognitive sub components. We then also discuss the limitations of our current approach.

9.1. Implications

As discussed in the introduction, the proposed model is based on the current understanding of the hierarchical structure and functional properties of the brain [43,56,66]. By proposing PASAR, our main goal was to propose an integrated framework that captures the interplay of the key cognitive sub components and to evaluate their contributions in real-world systems. Even though earlier proposals do not investigate the contributions of the individual cognitive sub components, our results have concrete implications for some of them.

Our model proposes top–down modulation of sensory data association by anticipating future stimuli. This has relevant implications for biological perception. Our model implies that stimulus anticipation results in a pretuning of local neural populations, which already finds some support in EEG studies [25]. Also in the concrete context of human vision, our results imply top–down modulation of bottom–up saccades and conscious decision making influenced by the anticipation of future stimuli. Supporting our hypothesis, a recent study in humans indicate that sensitivity of visual cortical neurons to sensory input is indeed dynamic during saccade preparation and dependent on the given attentional state [76]. Nevertheless, further experiments have to be carried out to validate the concrete implications of our model in biological perception.

A conceptual model of computation and representation in the brain was proposed recently that advocates the use of, so-called, Cortical Computational Units (CCU) respecting brain hierarchy for integrating attention, perception and motor action [4]. That proposal is a very strong candidate for a concrete instantiation of the computational mechanisms involved in such an integrated framework. Nevertheless, the individual contributions of attention, prediction, anticipation and memory on performance were not investigated. Furthermore, in contrast to our approach, no real-world instantiations of a CCU based cognitive framework are discussed.

A highly significant proposal for knowledge representation is Gardenfors' formulation of conceptual spaces, where *quality dimensions* which can be used to describe cognitive processes [27]. Recently an extension to this was proposed and demonstrated on a maritime application [64]. Current results on how data association is intricately involved in bottom–up sensory perception and the top–down bias on perception are in accordance with the conceptual space formulation for knowledge representation and emphasizes the role of *concepts* and their associations.

The recently proposed concept of the *cognitive niche* and the distribution of cognitive functions to the environment by humans, has an interesting relevance in our context [48]. It is argued that humans project ideas and thoughts onto external structures to aid cognitive performance. Nevertheless, here we argue that using predictions and anticipations to modulate sensory data association as proposed by PASAR provide yet another explanation based on the anatomical and functional understanding of the brain for optimizing cognitive resources, sparing the need for externalizations.

The extended Decision Field Theory (DFT) approach is a proposal for updating the subjective evaluation of available options and for attending on specific option attributes [45]. A Bayesian Belief Network (BBN) is applied to infer these updates in dynamic environments. Even though decision field theory is widely used in mathematical psychology, concrete support from neuroscience is yet to be found to validate the usage of DFT and BBN in the brain. Nevertheless, it would be of great interest to investigate the implications of the current results using DFT to see how subjective evaluations of options change with changing bottom–up sensory input and top–down anticipatory and attentional influences.

A neurodynamic model for look-ahead planning of sequential actions was proposed very recently, that was inspired by neurophysiological and behavioral findings from relevant studies of spatial behavior in primates [34]. Our model of integrating prediction, anticipation, sensation, attention and action provides a complementary approach to explaining look-ahead planning in animals. Besides that, PASAR also provides a concrete hypothesis on how such plans can be affected by top–down attentional or high-level knowledge.

The action selection and knowledge representation framework of the PASAR architecture can be formulated as an abstraction of the functionalities of the prefrontal cortex for task-oriented behavior selection and working memory, as proposed in other large-scale neurocomputational models [13,29]. Nevertheless, unlike the above models, PASAR integrates memory with attentional, predictive and anticipatory mechanisms in a single framework and concretely analyzes their individual contributions in real-world tasks.

We achieved robust performance results in real-world robotic tasks as opposed to most of the above theoretical models for integration of memory and behavior selection. Of high interest to the robotics and autonomous systems community is the fact that this will have implications for building novel real-world complex robotic cognitive systems, e.g. for humanoid platforms with high degrees of freedom. Our results on the contributions of transient memory, attentional, predictive and anticipatory mechanisms can thereby serve as a guiding principle for such real-world systems.

9.2. Limitations

Despite the concrete results achieved in real-world and simulated robotic tasks, the PASAR architecture needs more systematic tests in biological systems to actually prove the biological validity of the model. (See Section 10 to see what experiments are carried out at the moment and are planned for future to meet this challenge).

Furthermore, an interesting but open question in this context is the implication of our results for the so-called Mirror Neuron System (MNS) research [59,65]. A mirror neuron is a neuron that fires both when an animal performs an action and when the animal observes the same action performed by another individual. Possible connections of the MNS to perception–action coupling, intention prediction and theory-of-mind is still subject of ongoing research [12]. How the interplay of attention, prediction, anticipation, memory and actions as proposed by PASAR can benefit from the usage of an artificial MNS remains unclear. As the current proposal of PASAR does not explicitly contain mechanisms to explain *mirroring* of perception and actions, an extension of the framework would be needed to investigate this.

Besides, PASAR has to be tested in many more robotic implementations to further discover the strengths and weaknesses of the current proposal.

10. Conclusions and further research

We proposed PASAR, a concise and modular framework for evaluating quantitatively the integration of prediction, anticipation, sensation, attention and response, allowing the evaluation of system performance in given tasks using the whole system and also selected components of the system. We conducted experiments in simulations and complex real-world tasks to address the question of how each sub-component of PASAR contributes to the performance of the overall system in the given task. First we demonstrated the feasibility of our model for solving complex real-world problems using the XIM setup to solve a multi-person multi-modal tracking problem. In the XIM tracking task, we demonstrated that PASAR is able to learn sensory map alignment, create and maintain a world-model under limited resources constraints when acting in an unknown environment and deploy attentional mechanisms for active information acquisition. The second testbed of rescue robot simulation demonstrated PASAR's optimal usage of the acquired world-model for multiple goal decision making and allowed to evaluate the contributions of the individual components of PASAR to the overall performance. Current results on the interplay of anticipation, perception, attention and response suggest that a complete sensori-motor system with attentional, anticipatory and predictive mechanisms performs better than incomplete subsystems in complex tasks. We found evidence that the use of attentional mechanisms and prediction is only beneficial, if there is a forgetting (memory decay) mechanism at work. Building novel bio-mimetic robotic systems for real-world scenarios can benefit from our findings, by using prediction, anticipation, sensation, attention and response in the suggested combinations. We also believe that our results suggest more specific physiological and psychophysical experiments to support the findings of interplay of the different sub-systems of cognition in biological systems. It would be of great interest to perform psychophysical experiments to investigate the existence of anticipatory gates (validation gates) in human/animal perception.

The proposed model allows to evaluate the sub-components of an artificial cognitive system and we tested it in both simulations and real-world setups. Even though the model is based on the hierarchical architecture of biological systems and has been shown to be very efficient for artificial sensori-motor systems, concrete evidence of such a framework in biological systems is yet to be found. Targeted research has to be conducted to find support for the JPDA algorithm for data association in biological systems. Given the results of our model, we hypothesize the modulation of bottom–up saccades by high-level factors like the operational load. Some recent studies tend to indicate that sensitivity of visual cortical neurons to input is dynamic during saccade preparation, partly supporting our hypothesis [76]. Currently we are performing psychophysical experiments to test the hypothesis in human vision. Further psychophysical experiments are planned with neglect patients to investigate how spatial perceptual deficits can affect the anticipatory gate properties. Also we plan to investigate the existence of a neural correlate of the proposed anticipatory gate using EEG and fMRI studies in humans.

An important application of PASAR would be in real-world multi-robot scenarios, where a single monolithic robot is replaced by multiple simpler robots that have a single knowledge representation and predictive mechanisms, comparable to earlier proposals [21,47]. The question of how shared attention and predictive mechanisms can give rise to intelligent individual behavior patterns is of great interest in this context.

A very interesting and open research question is the usage of the modular concept of PASAR in the context of theory-of-mind to understand the interplay of attention, prediction, anticipation, memory and actions in humans [61]. Also in this context the connection to the Mirror Neuron System (MNS) and possible proposals for novel rehabilitation techniques for neglect patients based on our theoretical findings have to be investigated, similar to the recently proposed MNS based rehabilitation gaming system [12].

Also, we plan to test PASAR on the state-of-the-art 53 degree of freedom humanoid robotic platform iCub[8], to test its capability of efficient interaction in a multi-person interaction scenario. A humanoid robot platform offers a plethora of possibilities to test a framework like PASAR in real-world interactions with humans, which involves, prediction, anticipation, attention, sensation and motor response.

Another field of robotics research, where we plan to test PASAR, is autonomous landmark navigation. Landmark navigation in dynamic environments is a challenging problem as the navigator constantly has to anticipate the dynamics of the environment to successfully achieve navigation goals. Some recent proposals for light-weight autonomous landmark navigation in dynamic environments [52,53] could greatly benefit to achieve better results by making use of prediction and attention mechanisms for using limited navigational resources.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant agreement No. 216916 for NEUROCHEM, EU SYNTHETIC FORAGER (FP7-217148), PRESENCCIA (IST-2006-27731) and the Integrated CEEDS (ICT-258749) projects.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ins.2011.09.042.

References

[1] <http://opencv.willowgarage.com/wiki>.
[2] <www.bayesian-programming.org>.
[3] M. Ahissar, S. Hochstein, The reverse hierarchy theory of visual perceptual learning, Trends in Cognitive Sciences 8–10 (2004) 457–464.
[4] J.S. Albus, A model of computation and representation in the brain, Information Sciences 180 (2010) 1519–1554.
[5] J.R. Anderson, Act: a simple theory of complex cognition, American Psychologist 51 (1996) 355–365.
[6] A. Baddeley, Working memory, Science (1992) 556–559.
[7] Y. Bar-Shalom, T.E. Fortmann, Tracking and Data Association, Boston Academic Press, 1988.
[8] O. Berger, Statistical decision theory and bayesian analysis, Springer Series in Statistics (1980).
[9] U. Bernardet, M. Blanchard, P. Verschure, Iqr: A distributed system for real-time real-world neuronal simulation, Neurocomputing 44–46 (2002) 1043–1048.
[10] M. Botvinick, Z. Rosen, Anticipation of cognitive demand during decision-making, Psychological Research (2008) 1–8.
[11] R. Brooks, A robust layered control system for a mobile robot, IEEE Journal on Robotics and Automation 2 (1986) 14–23.
[12] M. Cameirão, S. Bermúdez, E. Oller, P. Verschure, Neurorehabilitation using the virtual reality based rehabilitation gaming system: methodology, design, psychometrics, usability and validation, Journal of NeuroEngineering and Rehabilitation 7 (1) (2010).
[13] G.L. Chadderdon, O. Sporns, A large-scale neurocomputational model of task-oriented behavior selection and working memory in prefrontal cortex, Journal of Cognitive Neuroscience 18 (2006).
[14] D.A. Chernyak, L.W. Stark, Top-down guided eye movements, IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics: A Publication of the IEEE Systems, Man, and Cybernetics Society (2001) 514–522.
[15] P. Cisek, J.F. Kalaska, Neural mechanisms for interacting with a world full of action choices, Annual Review of Neuroscience 33 (2010) 269–298.

[8] http://www.robotcub.org/.

[16] J. Collins, J. Uhlmann, Efficient gating in data association with multivariate gaussian distributed states, IEEE Transactions on Aerospace and Electronic Systems (1992) 909–916.
[17] P. Dean, J. Porrill, Adaptive-filter models of the cerebellum: computational analysis, Cerebellum 7 (4) (2008) 567–571.
[18] Y. Demiris, B. Khadhouri, Hierarchical attentive multiple models for execution and recognition (HAMMER), Robotics and Autonomous Systems 54 (2006) 361–369.
[19] P.F. Dominey, J.D. Boucher, Learning to talk about events from narrated video in the construction grammar framework, Artificial Intelligence 167 (2005).
[20] A. Duff, P.F.M.J. Verschure, Unifying perceptual and behavioral learning with a correlative subspace learning rule, Neurocomputing 73 (2010) 1818–1830.
[21] R.J. Duro, M. Graña, J. de Lope, On the potential contributions of hybrid intelligent approaches to multicomponent robotic system development, Information Sciences 180 (2010) 2635–2648.
[22] L.B. Ekstrom, P.R. Roelfsema, J.T. Arsenault, G. Bonmassar, W. Vanduffel, Bottom-up dependent gating of frontal signals in early visual cortex, Science 321 (2008) 414–417.
[23] J.G. Fleischer, J.L. Krichmar, Sensory integration and remapping in a model of the medial temporal lobe during maze navigation by a brain-based device, Journal of Integrative Neuroscience (2007) 403–431.
[24] K. Friston, S. Kiebel, Predictive coding under the free-energy principle, Philosophical Transactions of the Royal Society B: Biological Sciences 364 (2009) 1211–1221.
[25] I. Frund, J. Schadow, N.A. Busch, N. Naue, U. Korner, C.S. Herrmann, Anticipation of natural stimuli modulates EEG dynamics: physiology and simulation, Cognitive Neurodynamics 2 (2008) 89–100.
[26] J.M. Fuster, Inferotemporal units in selective visual attention and short-term memory, Journal of Neurophysiology 64 (3) (1990) 681–697.
[27] P. Gärdenfors, Conceptual Spaces: The Geometry of Thought, The MIT Press, 2000.
[28] J.I. Gold, M.N. Shadlen, Representation of a perceptual decision in developing oculomotor commands, Nature (2000) 390–394.
[29] M.E. Hasselmo, A model of prefrontal cortical mechanisms for goal-directed behavior, Journal of Cognitive Neuroscience 17 (2005) 1115–1129.
[30] H.R. Heekeren, S. Marrett, L.G. Ungerleider, The neural systems that mediate human perceptual decision making, Nature Reviews Neuroscience (2008) 467–479.
[31] M. Inderbitzin, S. Wierenga, A. Väljamäe, U. Bernardet, P. Verschure, Social cooperation and competition in the mixed reality space experience induction machine, Virtual Reality (2009) 153–158.
[32] L. Itti, P. Baldi, Bayesian surprise attracts human attention, Vision Research 49 (10) (2009) 1295–1306.
[33] L. Itti, C. Koch, Feature combination strategies for saliency-based visual attention systems, Journal of Electronic Imaging (2001).
[34] R. Ivey, D. Bullock, S. Grossberg, A neuromorphic model of spatial lookahead planning, Neural Networks 24 (3) (2011) 257–266.
[35] M.F. Jay, D.L. Sparks, Sensorimotor integration in the primate superior colliculus. I. Motor convergence, Journal of Neurophysiology (1987) 22–34.
[36] R.E. Kalman, A new approach to linear filtering and prediction problems, Transactions of the ASME–Journal of Basic Engineering 82 (Series D) (1960) 35–45.
[37] M. Kawato, Internal models for motor control and trajectory planning, Current Opinion in Neurobiology 9 (6) (1999).
[38] D. Knill, A. Pouget, The Bayesian brain: the role of uncertainty in neural coding and computation, TRENDS in Neurosciences 27 (2004) 12.
[39] E.I. Knudsen, Instructed learning in the auditory localization pathway of the barn owl, Nature (2002) 322–328.
[40] P. König, N. Krüger, Symbols as self-emergent entities in an optimization process of feature extraction and predictions, Biological Cybernetics (2006) 325–334.
[41] K.P. Kording, D.M. Wolpert, Bayesian integration in sensorimotor learning, Nature 427 (6971) (2004) 244–247.
[42] J.E. Laird, A. Newell, P.S. Rosenbloom, Soar: an architecture for general intelligence, Artificial intelligence 33 (1) (1987) 1–64.
[43] P. Langley, D. Choi, S. Rogers, Acquisition of hierarchical reactive skills in a unified cognitive architecture, Cognitive Systems Research 10 (2009) 316–332.
[44] P. Langley, K.B. McKusick, J.A. Allen, A design for the icarus architecture, SIGART Bulletin 2 (1991) 104–109.
[45] S. Lee, Y. Son, J. Jin, Decision field theory extensions for behavior modeling in dynamic environment using bayesian belief network, Information Sciences 178 (2008) 2297–2314.
[46] D. Liberati, L. Narici, A. Santoni, S. Cerutti, The dynamic behaviour of the evoked magnetoencephalogram detected through parametric identification, Journal of Biomedical Engineering 14 (1) (1992) 57–63.
[47] I. Lopez-Juarez, J. Corona-Castuera, M. Peña-Cabrera, K. Ordaz-Hernandez, On the design of intelligent robotic agents for assembly, Information Sciences 171 (2005) 377–402.
[48] L. Magnani, Abducing chances in hybrid humans as decision makers, Information Sciences 179 (2009) 1628–1638.
[49] P.C. Mahalanobis, On the generalised distance in statistics, Proceedings National Institute of Science, India 2 (1) (1936) 49–55.
[50] Z. Mathews, S. Bermúdez, P. Verschure, A novel brain-based approach for multi-modal multi-target tracking in a mixed reality space, in: Intuition Workshop and Conference on Virtual Reality and Virtual Environments, 2007.
[51] Z. Mathews, S. Bermúdez, P. Verschure, Intelligent motor decision: From selective attention to a bayesian world model, in: IEEE Intelligent Systems Conference, 2008.
[52] Z. Mathews, S. Bermúdez, P. Verschure, An insect-based method for learning landmark reliability using expectation reinforcement in dynamic environments, in: Robotics and Automation (ICRA), IEEE/RSJ International Conference, 2010.
[53] Z. Mathews, M. Lechón, J.M. Blanco, A. Dhir, A. Duff, S. Bermúdez, P. Verschure, Insect-like mapless navigation based on head direction cells and contextual learning using chemo-visual sensors, Intelligent Robots and Systems (IROS), in: IEEE/RSJ International Conference, 2010.
[54] J.S. Nairne, Remembering over the short-term: the case against the standard model, Annual Review of Psychology 53 (2002) 53–81.
[55] L. Narici, D. Liberati, S. Cerutti, A. Santoni, Analysis of the neuromagnetic data on the frequency responsivity of the human brain: a progress report, Clinical Physics and Physiological Measurement 12 (A) (1991) 43.
[56] H. Nienborg, B.G. Cumming, Decision-related activity in sensory neurons reflects more than a neuron's causal effect, Nature (2009) 89–92.
[57] S. Oh, S. Sastry, A polynomial-time approximation algorithm for joint probabilistic data association, in: American Control Conference, 2005 1283–1288.
[58] B.A. Olshausen, C.H. Anderson, D.C.V. Essen, A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information, The Journal of neuroscience: the Official Journal of the Society for Neuroscience 13 (11) (1993) 4700–4719.
[59] G.D. Pellegrino, L. Fadiga, L. Fogassi, V. Gallese, G. Rizzolatti, Understanding motor events: a neurophysiological study, Journal of Biomedical Engineering 91 (1) (1992) 176–180.
[60] M.A. Pinsk, G. Doniger, S. Kastner, Push-pull mechanism of selective attention in human extrastriate cortex, Journal of Neurophysiology 92 (2004) 622–629.
[61] D.G. Premack, G. Woodruff, Does the chimpanzee have a theory of mind?, Behavioral and Brain Sciences 1 (1978) 515–526
[62] T.J. Prescott, P. Redgrave, K. Gurney, Layered control architectures in robots and vertebrates, Adaptive Behavior 7 (1999) 99–127.
[63] R.P. Rao, D.H. Ballard, Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, Nature Neuroscience 2 (1) (1999) 79–87.
[64] J.T. Rickard, J. Aisbett, G. Gibbon, Reformulation of the theory of conceptual spaces, Information Sciences 177 (2007) 4539–4565.
[65] G. Rizzolatti, L. Fadiga, V. Gallese, L. Fogassi, Premotor cortex and the recognition of motor actions, Cognitive Brain Research 3 (2) (1996) 131–141.
[66] B.E. Stein, The development of a dialogue between cortex and midbrain to integrate multisensory information, Experimental Brain Research 166 (2005) 305–315.
[67] B.E. Stein, M.A. Meredith, The Merging of the Senses, The MIT Press Cambridge, Massachusetts, London, England, 1993.

[68] S. Treue, Neural correlates of attention in primate visual cortex, Trends in Neurosciences 24 (5) (2001) 295–300.
[69] S. Ullman, Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex, Cerebral Cortex 5 (1) (1995) 1–11.
[70] P. Verschure, P. Althaus, A real-world rational agent: unifying old and new ai, Cognitive Science A Multidisciplinary Journal 27 (4) (2003) 561–590.
[71] P. Verschure, B.J.A. Kröse, Rolf Pfeifer, Distributed adaptive control: the self-organization of structured behavior, Robotics and Autonomous Systems 9 (3) (1992) 181–196.
[72] P. Verschure, T. Voegtlin, R.J. Douglas, Environmentally mediated synergy between perception and behaviour in mobile robots, Nature 425 (6958) (2003) 620–624.
[73] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, Computer Vision and Pattern Recognition 1 (2001) 511–518.
[74] I.B. Witten, E.I. Knudsen, H. Sompolinsky, A hebbian learning rule mediates asymmetric plasticity in aligning sensory representations, Journal of Neurophysiology 100 (2) (2008) 1067–1079.
[75] S. Wood, Representation and purposeful autonomous agents, Robotics and Autonomous Systems Knowledge Engineering and Ontologies for Autonomous Systems, AAAI Spring Symposium 49 (1–2) (2004) 79–90.
[76] H. Xue, S.X. Xian, T. Moore, Dynamic sensitivity of area V4 neurons during saccade preparation, Proceedings of the National Academy of Sciences 106 (31) (2009) 13046–13051.