2018-06

Visual attention and object naming in humanoid robots using a bio-inspired spiking neural network

Hernandez Garcia, D

http://hdl.handle.net/10026.1/11493

10.1016/j.robot.2018.02.010
Robotics and Autonomous Systems
Elsevier

Visual attention and object naming in humanoid robots using a bio-inspired spiking neural network

Article *in* Robotics and Autonomous Systems · March 2018

DOI: 10.1016/j.robot.2018.02.010

6 authors, including:

Daniel Hernandez Garcia
University of Plymouth
13 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Samantha Vanessa Adams
University of Plymouth
20 PUBLICATIONS **26** CITATIONS
SEE PROFILE

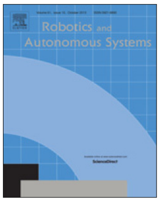
Thomas Wennekers
University of Plymouth
66 PUBLICATIONS **471** CITATIONS
SEE PROFILE

Angelo Cangelosi
University of Plymouth
293 PUBLICATIONS **4,006** CITATIONS
SEE PROFILE

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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



Visual attention and object naming in humanoid robots using a bio-inspired spiking neural network

Daniel Hernández García[a,*], Samantha Adams[a], Alex Rast[b], Thomas Wennekers[a], Steve Furber[b], Angelo Cangelosi[a]

[a] *Centre for Robotics and Neural Systems, Plymouth University, Plymouth, UK*
[b] *School of Computer Science, Manchester University, Manchester, UK*

HIGHLIGHTS

- Develop a neuroanatomically grounded spiking neural network for visual attention with a word learning capability.
- Demonstrates that a label could be associated with a salient object via Spike-Timing Dependent Plasticity in a simple system.
- Provides a proof-of-concept case for the integration of biologically inspired neural networks with robotics for basic language acquisition.

ARTICLE INFO

Article history:
Received 3 May 2017
Received in revised form 21 November 2017
Accepted 18 February 2018
Available online 5 March 2018

Keywords:
Neurorobotics
Object naming
Visual attention
Biological inspired models
Spiking neural networks

ABSTRACT

Recent advances in behavioural and computational neuroscience, cognitive robotics, and in the hardware implementation of large-scale neural networks, provide the opportunity for an accelerated understanding of brain functions and for the design of interactive robotic systems based on brain-inspired control systems. This is especially the case in the domain of action and language learning, given the significant scientific and technological developments in this field. In this work we describe how a neuroanatomically grounded spiking neural network for visual attention has been extended with a word learning capability and integrated with the iCub humanoid robot to demonstrate attention-led object naming. Experiments were carried out with both a simulated and a real iCub robot platform with successful results. The iCub robot is capable of associating a label to an object with a 'preferred' orientation when visual and word stimuli are presented concurrently in the scene, as well as attending to said object, thus naming it. After learning is complete, the name of the object can be recalled successfully when only the visual input is present, even when the object has been moved from its original position or when other objects are present as distractors.

1. Introduction

Current research in behavioural and cognitive neuroscience demonstrates a close link between the brain systems for language, action and perception [1–6]. Advances in behavioural and computational neuroscience and cognitive robotics provide a timely opportunity to integrate the interdisciplinary methods and approaches from these fields with the aim of furthering the scientific and technological progress in language processing and embodied artificial cognitive systems. In the embodied approach to language acquisition, auditory input, visual input and motor interaction with the world are considered equally important [7]. Neuroscience experiments show that the use of language activates "embodied representations", that is, brain areas and circuits closely joining together motor, perceptual and speech-language mechanisms [8]. This finding is consistent with the embodied view of cognition in psycholinguistics and cognitive science where cognitive functions, such as language, are intertwined with sensorimotor knowledge [9], and with the situated learning approach to studying language in context. Research in behavioural and cognitive neuroscience has demonstrated that language, action and perception are closely linked in the brain [10,7,11].

Developments in neuroscience and cognitive research have been closely followed by advances in the design of neuroanatomically grounded models of word acquisition [12,13] and by the use of such brain-inspired models in the integration of action and language learning in robots [14,15], in visuo-motor integration [16] and in visual attention [17]. For example, Garagnani et al. [13] designed multi-layer neural networks whose architecture is neuroanatomically grounded in the left perisylvian language cortex.

[*] Corresponding author.
E-mail address: daniel.hernandez@plymouth.ac.uk (D. Hernández García).

https://doi.org/10.1016/j.robot.2018.02.010

Simulations reproduced the cortical responses to familiar versus pseudo-word stimuli. Cognitive robotics has previously made use of brain-inspired models in embodied contexts. Morse et al. [14] trained the humanoid robot iCub to learn the names of objects, replicating similar phenomena observed in child language experiments. Caligiore et al. [15] developed the TRoPICALS model to study how vision, action and language are integrated in the representation and activation of affordances. Adams et al. [16] used a neuroanatomically grounded neural model with the iCub humanoid robot simulator to explore learning of associations between visual and motor modalities. Adams et al. [17] integrated a spiking neural model for featured based attentional selection with the iCub to enable the robot to perform a behavioural task: fixating attention upon a selected stimulus, this is enacted by directing the robot's gaze towards the stimulus based on the spiking activity of the neural model.

In the current work the visual attention model from Adams et al. [17] has been extended to add language learning capabilities. In the original attention network objects are salient and attended to depending upon their shape (orientation). With the addition of an auditory modality in the present work, words are presented when the salient object is attended to and, through learning, the word label is associated with the object. Learning of the label is independent of the object's position such that when the auditory stimulus is removed the label can still be recalled on presentation of the visual stimulus even when the object was moved from its original position.

Yu and Smith recently found a strong link between object naming and visual attention in experiments with children [18]: object naming only successfully took place when the object was fully attended to — i.e. centred in the child's field of view and dominating the scene. The word learning processes in the model are based entirely on mechanisms of Hebbian plasticity [19]. In the current model, learning the association between the auditory and visual modalities has been implemented using Spike-Timing Dependent Plasticity (STDP), a mathematical formulation for modelling learning in real neurons [20]. The model developed during the current work provides a basis for a future developmental robotics approach to language learning, as attention to and naming of individual objects forms the first stage of lexical development [21]. This, in addition to more complex visual and motor modalities, is required for the development of higher cognitive abilities such as reasoning about objects and tasks.

In our model objects can be biased according to their shape, in this case the shape is defined by the object's orientation. Following the nomenclature provided in [22,17], objects will be designated as 'preferred' when positively biased, 'aversive' or 'non-preferred' when negatively biased, and 'neutral' or 'unbiased'. Preferred objects are attended to by the robot looking at them, changing its gaze to fix the object in the centre of its visual field, while aversive and neutral objects are ignored and provide not activation from the attention network. The model (see [17] for the original attention model) has the capability to learn which types of objects are preferred but here we have used hardcoded preferences in order to focus on the object naming mechanism. Therefore, before the learning of an object's name takes place preferred objects are already recognized as salient and are attended to but are 'unlabelled'. Results show that with a simple extension to add an auditory modality and learning using STDP, a label can be associated with the preferred object via its orientation such that when the auditory stimulus is removed the label can still be recalled on presentation of the visual stimulus.

The structure of this paper is as follows. A short background on previous works in neurobiologically inspired robotics (Section 2) is given first, followed by a description of the visual attention neural network and of the extensions made to enable word learning (Section 3). Next the iCub robot and the experimental platform are introduced (Section 4). Finally, we describe the object naming experiments and present our results (Section 5) and conclusions (Section 6).

2. Background on neurobiologically inspired robotics

Neurorobotics is not encapsulated in a single field it ranges across many disciplines such as computer science, engineering, neuroscience, and others. The field is based on the embodied approach to cognition. The grand challenge of neurorobotics is to build a well-founded experimental science of embodiment [23]. Because the nervous system is so closely coupled with the body and situated in the environment, brain-based robots can provide powerful tools for studying neural function, as they can be tested and probed in ways that are not yet achievable in human and animal experiments. Neurobiologically inspired systems present great potential to advance the field of autonomous robots and our understanding of the human brain [24].

Experiments in neurorobotics permit to progress in understanding how the interplay between neural learning dynamics, physical embodiment and environmental factors shape developmental trajectories [23]. While relatively few studies can be found addressing the implementation of biologically inspired robot designs and neural architectures that lead to brain-based robots [25], we can highlight a small number of relevant and similar approaches in the development of neuromorphic cognition. Rucci et al. [26] provide examples of robotic systems work in the areas of sensory perception and motor learning. In Krichmar and Cox [27] a strategy for controlling autonomous robots based on the principles of neuromodulation in the mammalian brain is presented. Galluppi et al. [22] provides the implementation of a neural network for feature-based attention integrating a visual AER sensor and the SpiNNaker system. Adams et al. [17] integrated a spiking neural model for visual attention with the iCub to enable the robot to perform a behavioural task: fixating attention upon a selected stimulus. The work of de Azambuja and colleagues [28] use Liquid State Machines (LSM) to learn trajectories with the BAXTER robot. Gamez et al. [29] present a spiking neural interface for the iCub robot, "iSpike". Barros et al. [30] present a model that uses a hierarchical feature representation to deal with spontaneous emotions, and learns how to integrate multiple modalities for non-verbal emotion recognition, making it suitable to be used in an HRI scenario. Park and Tani's work [31] presents neurorobotics experiments on acquiring skills for "communicable congruence" with humans via learning. Seepanomwan et al. [32] propose a novel neurorobotic model that has a macro-architecture constrained by knowledge held on the brain, encompasses a rather general mental rotation mechanism, and incorporates a biologically plausible decision making mechanism. In Beyeler et al. [33] a cortical neural network model for visually guided navigation has been embodied on a physical robot exploring a real-world environment. The work of Walter et al. [34] provides an overview of available neuromorphic chip designs and analyse them in terms of neural computation, communication systems and software infrastructure, as well as review neurobiological learning techniques.

A commonality among neurorobotic approaches is that they are neuromorphic in their architecture; they contain neuronal elements and synaptic connectivity inspired by what is currently known about the nervous system; and they are embedded on physical devices [24]. A wide variety of computational approaches can be used to control neurobiologically inspired robots, including spiking neural networks, firing rate neurons, recurrent neural networks, and dynamic neural fields [22,17,28,29,31].

Studies attempting to use Spiking Neural Networks (SNN) for practical applications demonstrate promising results in solving complex real world problems. SNNs seem to be able to solve difficult cognitive problems [35] in possibly nonstationary environments [36]. Perhaps spike based neurorobots can embody behavioural features that are difficult or impossible using other methods [37]. An attractive factor of the approach is its potential to inform the neurosciences as well as the robotic domain [34].

3. Biologically inspired neuroanatomical model of visual attention and language grounding

There is strong neuroscientific evidence showing that activation of brain areas, responsible for motor, perceptual and speech-language mechanisms relates to the use of language [8]. These results are consistent with the psycholinguistics and cognitive science embodied view of cognition, for which cognitive functions, such as language, are closely integrated with sensorimotor knowledge [9]. Various connectionist models of word learning and language processing exist [38–43]. While these models have provided important contributions to the understanding of how different parts of the human brain may play an active role in language processing, in general they fall short of providing a mechanistic explanation of the neurobiological mechanisms at the basis of language acquisition and processing, and at their neurobiological plausibility [5].

The basis for the model used in this work is an attentional model, inspired by the primate visual system, first described in [44]. This network was subsequently reformulated as a spiking neural network and adapted to run on the SpiNNaker platform [22]. In a previous work [17], the network was adapted to generate visual attention behaviour for the iCub robot.

In the present work an auditory modality and learning using Spike-Timing Dependent Plasticity (STDP) was added to extend the visual attention network in [17] with multi-modal and auditory areas for a word learning capability. In the attention network objects are salient and attended to depending upon their orientation. The addition of an auditory modality allows a word label to be associated with the salient object attended to through learning. Learning is position independent such that when the auditory stimulus is removed the label can still be recalled on presentation of the visual stimulus even when the object was moved from its original position. The present work was implemented using the Python PyNN interface language for SNNs but future works will be directed to implemented in the SpiNNaker neuromorphic platform. The network should transfer seamlessly to SpiNNaker given that previous work using only the visual attention portion of the network has already done so [17].

3.1. The spiking neuron model

Spiking neuron models process information coming from many inputs to produce single spiking output signals. A SNN is supposed to generate one or more spikes, when internal variables of the model reach a certain state, with a probability increased by excitatory inputs and decreased by inhibitory inputs [45]. A neuron fires whenever its "potential", the sum of excitatory postsynaptic potentials ("EPSPs") and inhibitory postsynaptic potentials ("IPSPs"), reaches a certain threshold Θ. Postsynaptic potentials result from the firing of other neurons connected through "synapses". The firing of a "presynaptic" neuron u at time s contributes to the potential P_v of the spiking neuron v at time t an amount that is modelled by the term $w_{u,v} \cdot e_{u,v}(t-s)$, which consists of a "weight" $w_{u,v} \geq 0$ and a response function $e_{u,v}(t-s)$.

The most widely used and most common model of spiking neurons is the class of integrate and fire models, the Integrate-and-Fire (IF) and Leaky-Integrate-and-Fire (LIF) models are the best-known examples of formal spiking neuron models [46]. Both of these models treat biological neurons as point dynamical systems, and neglect the spatial structure properties of biological neurons [45]. While these are still simplified models, focusing on just a few aspects of biological neurons, they are substantially more realistic in comparison with previous neural models [47]. For the IF and the LIF neuron the shape of the action potentials is neglected and every spike is considered as a uniform event fully characterized by the time of its appearance [48]. The model is based on the principles of electronic circuits. The LIF basic circuit consists of a capacitor C in parallel with a resistor R driven by a current $I(t)$. The membrane potential in the LIF neuron is described by the single first-order linear differential equation $\tau_m \frac{du}{dt} = -u_{rest}(t) + RI(t)$, where $\tau_m = RC$ is taken as the time constant of the neuron membrane, modelling the voltage leakage, and u as the membrane potential. A spike firing time $t^{(f)}$ is defined by a threshold criterion $u(t^{(f)}) = \vartheta$. Immediately after $t^{(f)}$, the potential is reset to a new value $u_{rest} < \vartheta$. An absolute refractory period can be modelled by forcing the neuron to a value $u = u_{abs}$ during an absolute refractory time Δ_{abs} after a spike emission, and then restart the integration at time $t^{(f)} + \Delta_{abs}$, with initial value $u = u_{rest}$. The combination of leaky integration and reset defines the basics of the Leaky-Integrate-and-Fire model [46]. This model is computationally simple and can be implemented in hardware like the SpiNNaker system [49].

Spike-Timing Dependent Plasticity (STDP) is a form of competitive *'Hebbian learning'* that uses exact spike timing information [20,50]. Experimental and modelling studies have shown that this form of 'Hebbian' plasticity, where the relative firing times of pre and postsynaptic neurons influence the strengthening or weakening of connections, covers central aspects of the mechanism that real neurons use [20]. When the presynaptic spike is emitted before the postsynaptic spike there is potentially a causal relationship and the connection is strengthened, long-term potentiation (LTP). When firing times cannot be causally related (i.e. the postsynaptic spike occurs before the presynaptic one) then the synapse is weakened, long-term depression (LTD). Postsynaptic neurons are sensitive to the timing of incoming presynaptic potentials, which leads to competition among the presynaptic neurons. This sensitivity can result in shorter latencies, spike synchronization and faster information propagation through the network [50,20].

The STDP learning rule is defined by a function $F(\Delta t)$ which determines the amount of synaptic modification, weight change, dependence on a single pair of pre- and postsynaptic spikes separated by a time Δt,

$$F(\Delta t) = \begin{cases} A_+ \cdot exp(\Delta t/\tau_+) & \text{if} \quad \Delta t < 0 \\ -A_- \cdot exp(-\Delta t/\tau_-) & \text{if} \quad \Delta t \geq 0 \end{cases} \tag{1}$$

where, τ_+ and τ_- are time constants that determine the ranges of pre-to-postsynaptic interspike intervals over which LTP and LTD occur. A_+ and A_- are positive amplitudes determining the maximum amounts of synaptic modification, which occur when Δt is close to zero [20]. Connection weights are hard limited to lie between 0 and a upper maximum value, *gmax*. The function $F(\Delta t)$ would result in weight changes for LTP and LTD when Δt is < 0 or ≥ 0 respectively. The change of the peak conductance, g, at a synapse due to a presynaptic spike occurring at time t_{pre} and a postsynaptic spike at time t_{post} is given by $g \longrightarrow g + gmax * F(\Delta t)$, where $\Delta t = t_{pre} - t_{post}$. If this modification would make g less than 0 or greater than *gmax*, g is set to the appropriate limiting value [51].

3.2. Spiking neural network of visual attention and object naming

Fig. 1 shows the visual attention model and the object naming extension to the model to add multi-modal and auditory areas. The original attention model architecture of [22,17] consists of six

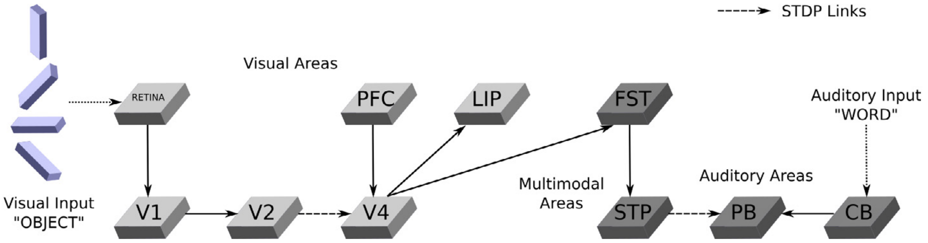


Fig. 1. Neural model architecture with relevant brain areas and connectivity. Original attention model architecture from [17] in lighter grey, object naming extension to the model with auditory and multi-modal areas shown in a darker grey. Arrow heads indicate direction of connections, dashed links indicate STDP is enabled in the connection. V1–V2–V4 — Visual Cortex areas 1–2–4, PFC — Prefrontal Cortex, LIP — Lateral IntraParietal area, FST – Fundus of the Superior Temporal Sulcus, STP – Superior Temporal Polysensory area, PB – Auditory Parabelt area, CB – Core and Belt combined auditory area.

areas that model a specific set of cortical areas in the mammalian eye and brain. The Retina area corresponds roughly with the retina and Lateral Geniculate Nucleus (LGN) in the real visual system and consists of layers of 'ON' and 'OFF' cells. Areas V1, V2 and V4 correspond to known visual processing areas in the occipital lobe of mammalian cortex. The PFC area corresponds to Prefrontal Cortex. The LIP area corresponds to the Lateral IntraParietal cortex. All areas are topographically mapped to the input space so that a neuron represents a fixed visual position in the input image. The model was extended, with a biologically inspired architecture, to enable the association of a label with an object adding two multi-modal and two auditory areas. The CB area combines the auditory Core and Belt. The PB area corresponds to the auditory Parabelt. Areas FST and STP represent the Fundus of the Superior Temporal Sulcus and Superior Temporal Polysensory respectively of the Superior Temporal Sulcus (STS). The Retina represents a 32 × 32 neuron visual field, while V1 and V2 are 20 × 20 image maps (400 neurons), and V4, PFC, FST, STP, PB and CB are 10 × 10 neuron matrixes; in total the systems simulates 6124 neurons. Each one of the areas V1, V2, V4, PFC, FST and STP are separated into 4 orientation-specific layers (Horizontal: —, Vertical: |, Diagonal: \and Counterdiagonal: /). The LIP area merges the 4 orientation layers via a winner-take-all. The CB area process auditory input. PB forms the link to the multimodal areas in order to make associations with the visual modality.

In the model, area V1's role is orientation selectivity. The V1 area provides functionality in a similar manner to cortical area V1 in the human brain, which contains neurons selective to stimulus orientation [52]. V1 layers contain tunable orientated Gaussian filters, equally dividing the possible angles in $[0, \pi)$ radians, that perform low-level orientation discrimination. The spike input from the Retina is connected to area V1 implementing a convolutional network with different Gaussian orientation filters. Each neuron in area V1 is one to one connected to the equivalent neuron in area V2 for each of the 4 orientation layers.

Area V2 is a pooling and competition layer between the 4 orientations. V2 is also tuned for orientation but at a coarser resolution than V1. Area V2 pools input from a local neighbourhood of V1 neurons to merge patches into orientated edges; subsampling the activity of neurons with the same preferred orientation is implemented by a localmax function, in a manner similar V2 complex cells [53]. Each V2 sublayer has internal lateral inhibition to form a soft-winner-take-all and a global winner-take-all among the sublayers. The local competition between neurons with different preferred orientations sustained the activity of neurons whose preferred orientation matched the stimulus, and suppressed activity relating to non-matching neuronal responses [22]. The V2 to V4 connection has STDP enabled which results in faster saccades and more persistent attention to preferred stimuli.

PFC area functions in the current model as a memory area for hardcoding preferred orientation of an object. In real brains, PFC is involved in many functions related to complex cognitive behaviour and has been implicated in remembering object locations during selective attention [54]. Four neuronal populations in the PFC (memory) area encode the goal of selecting a stimulus with a particular orientation. Each V4 layer receives (initially hardwired) bias from a PFC layer which determines a top-down source of preference. In the model, the orientation of objects can be designated as 'preferred', 'non-preferred' or 'neutral'. This is implemented using fixed biases applied via the PFC area. The initial and resting membrane potentials of the PFC neurons are randomly initialized with a uniform distribution and the offset current is set to hardwire the preference in orientation. The model has the capability to learn which types of object's orientations are preferred but in this work hardcoded preferences have been used in order to focus on the object naming mechanism.

Area V4 is a biasing layer; is also tuned for orientation, groups lower level features into shapes and is also subject to attentional modulation. V4 area neurons are analogous to the neurons of cortical area V4, which receives a large input from working memory via the frontal eye field [55]. Area V4 receives combined activity from V2 (pooling and competition) and PFC (memory) layers, such that activity is maximized for stimuli of the desired orientation, provided they are also present in the visual field [22]. V4 groups edges into objects by locally subsampling V2 neurons in their matching sublayer.

The LIP area works as a selection layer to form a retinotopic visual saliency map. Experiments have shown that neurons in LIP store location information that guides movement to fixate upon a target [56]. The LIP area performs object selection via a winner-take-all mechanism and identifies the salient location of the object with the preferred orientation in the scene, the most active neuron from the LIP layer provides the location output, saliency spikes 3 and its transformed into a point in the robot visual field, so that the iCub robot can look at the object. The Robot moves using the iKinGazeCtrl [57] to attend to the salient object, this includes rotations of the head and neck, and movement of the eyes. V4 neurons project one-to-one to a merged-orientation LIP layer which provides a first stage of output selection. The LIP also includes a hard winner-take-all pattern of lateral inhibition to select a single attentional position at each moment, driving actuators which direct motion. Activity at this location (i.e. the target of attention) was maintained, while activity at other locations was suppressed [22].

In our model, area CB represents the first areas of the auditory cortex. In the brain, the auditory system is subdivided into three areas, A1 primary auditory cortex, auditory Belt and Parabelt [58]. The CB area combines both the auditory Core and auditory Belt in our model. The auditory Core, including primary auditory cortex (A1), and auditory Belt are topographically connected with a tonotopic arrangement [59]. Justification for combining auditory Core and auditory Belt comes from assuming the Belt is a coarser tonotopic representation of the Auditory Core and so for the purpose of modelling there is no benefit to separating these two areas unless one is modelling a hierarchical tonotopy to process actual auditory input [59,58].

The PB area represents the second auditory area, the Parabelt, and forms the link to the multi-modal areas in order to make associations with the visual modality. In the auditory system, the Parabelt lies within a third level of cortical processing in a core–belt–parabelt pathway [58,59]. The auditory processing is extended beyond auditory cortex via connections of the Parabelt with specific regions of adjacent temporal cortex, medial temporal cortex, prefrontal cortex, and parietal cortex [58]. The connections from the PB area are used to link visual and auditory areas in the model via a multi-modal area modelled in the Superior Temporal Sulcus, since there is evidence that the Parabelt connects to polysensory areas with nearby cortex of the upper and lower banks of the Superior Temporal Sulcus [59].

In the model FST and STP areas represent the multi-modal area of the Superior Temporal Sulcus (STS); here subdivided into caudal and rostral areas, namely the Fundus of the Superior Temporal Sulcus (FST) and Superior Temporal Polysensory (STP). The caudal parts of STS are known to be occupied by visual areas, but there is no evidence for direct Parabelt connections with these visual areas. However, the more rostral parts of the STS appear to be polysensory, with neurons responding to auditory, visual, and even somatosensory stimulation [59,60]. In this model the V4 visual area connects to FST as there is anatomical evidence for such a connection in primates [61]. In our model learning the association of a label with an object takes place on the bridge between the auditory and multi-modal areas, that is, the PB to STP connection which are in the model random, sparse and enabled with STDP. As STP–PB and LIP are not themselves directly connected but indirectly associated by their connections with V4, this in theory means that object naming should be location independent. In our experiments, the visual field of the robot remains static for the duration of the learning and recalling phase. This simplifies the problem as the visual shape of the object, its orientation will not be changing due to the movement and rotation of the robots head, creating additional problems that are not addressed in this paper.

3.3. Non bio-inspired approaches for language grounding learning

Symbol grounding and embodied language learning have been an attractive research topic for cognitive and developmental robotics [62,63]. These issue have been addressed by many authors and there have been different approaches and several grounded language learning architectures proposed in the literature. For example, the work of Saunders and his collaborators focus on grounding lexical concepts in a robot's sensorimotor activity via human–robot interaction [64,65]. In [66] prosodic analysis and extraction of salient words are associated with a robots sensorimotor perceptions for acquisition of lexical meaning, in an attempt to ground these words in the robots own embodied sensorimotor experience. Lyon et al. [67,68] provides a developmental robotics model of the transition from babbling to word forms with the iCub robot. Their work demonstrates a platform in which it is possible to sustain interaction to achieve rudimentary word form acquisition in real-time using a simple frequency dependent probabilistic generation mechanism, together with human reinforcement [68]. The work Föerster et al. [69] addresses the acquisition of the word "no" and of the concept of negation. Their cognitive architecture extends symbol grounding beyond the realm of sensorimotor-data to encompass affect, in their system the utterance of negation words is grounded in a negative affective/motivation state and often have prosodic saliency [69]. This taxonomy have informed recent developmental robotics studies of the role of affective behaviour in the acquisition of negation. Steels and his group, [70,71], have used hybrid population of robots, Internet agents, and humans engaged in language games. In their work relevance is given to the social aspects of the symbol grounding, as well as the perceptual grounding of categories [72]. The work of Nakamura and others, [73,74], deal with multimodal sensory information by using a latent Dirichlet allocation (LDA)-based framework for multimodal categorization and words grounding by robots. [75] shows multimodal categorization based on the autonomously acquired multimodal information and partial words given by human users and [74] proposed an unsupervised method to generate natural sentences from observed multimodal information in a bottom up manner using multilayered multimodal latent Dirichlet allocation and Bayesian hidden Markov models. Many others examples can be found, see [76–79] for a review of robotics models of the grounding of language.

It was not the motivation of this work to provide a comparison between cognitive and developmental, non-bio-inspired, approaches to language learning and grounding but to highlight how a neuroanatomically grounded model for visual attention can be extended with a word learning capability and validate its real-time implementation with the iCub humanoid robot to demonstrate attention-led object naming. Here we aim at more realistic modelling of spiking neuron activity in large neuronal assemblies (6124 total neurons in the implementation reported in this paper and scalable up to tens of thousands of neurons in future work) distributed across a range of cortical areas.

Integrating spiking neural networks with robots introduces considerable complexity yet providing no significant benefit in task performance, where non bio-inspired robotic solutions or abstract neural simulations can usually produce better-performing and more informative results. But, we suggest, in a cognitive robotics context, where the goal is understanding computations of the brain, such an approach may yield useful insights to neural architecture as well as learned behaviour. Advances in understanding the neurobiology suggest that neural models more closely matching the biology can help reveal the computational principles necessary for cognitive robotics while illuminating human brain function [80]. One of the principal contributions of the cognitive neurorobotics approach is that it allows to pursue both the study of the neuroscience of the brain and the engineering of functional robots in the same context as a tool to uncover the model of computation, and then in a recursive process take the insights thus gained to refine the model systematically and produce systems that function in the real world. Next stage of our research will take advantage of the huge benefits offered by advances in neuromorphic hardware systems, since development of neuromorphic hardware [81] and of the robotic systems have now reached a point of maturity where integrated neurorobots able to demonstrate effective behaviour in nontrivial real-world scenarios are within reach [17]. This type of model will lead to more precise predictions and stronger experimental validation of the theory.

4. Experimental platform

Fig. 2, shows the experimental setup for the object naming experiments for the simulated (left) and real (right) iCub humanoid robot. Only two objects of different orientations (horizontal and vertical) are ever present in the scene. One of these orientations (horizontal or vertical) will be biased to be 'preferred' while the other will be biased as 'non-preferred' or 'aversive', diagonal and counterdiagonal are 'unbiased' or 'neutral'. Therefore, before the learning of an object's name takes place preferred objects are already recognized as salient and are attended to but are 'unlabelled'.

4.1. iCub humanoid robot

The iCub simulator is an accurate physical simulation of the real iCub (Fig. 2) that can be used to develop applications that are easily transferable to the real robot with minimal changes.

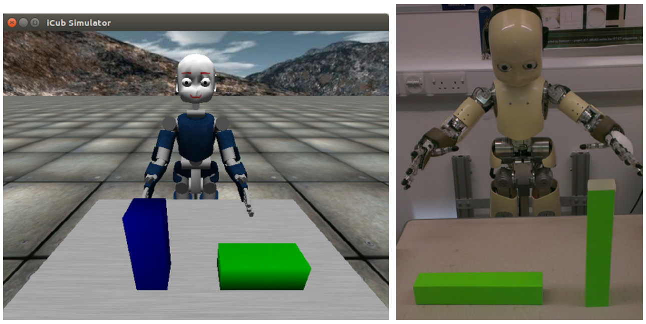


Fig. 2. The iCub simulated environment (left) and the real iCub (right).

Table 1
LIF and STDP parameters for the neuron model.

Parameter	Description	Value
V_{rest}	Resting potential	−65 mV
V_{reset}	Reset potential	−65 mV
V_{thresh}	Threshold potential for spiking	−45 mV
T_{refrac}	Refractory period	3.0 ms
$\mathcal{T}_m$	Membrane time constant	24.0 ms
$\mathcal{T}_+$	LTP time constant	20.0 ms
$\mathcal{T}_-$	LTD time constant	20.0 ms
gmax	Maximum synaptic weight	20.0 nA
$\mathcal{A}_+$	LTP weight update amplitude	0.05*gmax*
$\mathcal{A}_-$	LTD weight update amplitude	0.0675*gmax*

As well as basic motor control and visual processing, for more specialized tasks the iCub simulator (and real robot) can be integrated with external libraries such as the OpenCV image processing framework. For communications, iCub uses YARP, a generic and flexible protocol which we used to connect to the neural model. Fig. 3 gives an overview of the integrated system which we used for both simulated and real iCub. For convenience, we also used Aquila, a software architecture for cognitive robotics designed to provide useful functionality for iCub applications [82]. In particular we used the Tracker module for the extraction of objects from the scene and the iCubMotor module to convert image coordinates into head motor movements to enable the iCub to look at a location corresponding to a point in a 2D image and also point towards it.

4.2. Spiking neuron network

The model used in the current work was described in Section 3.2. We implemented a network consisting of 10 areas of artificial spiking neurons modelling regions of the visual and auditory cortex. Fig. 1 shows the Visual Attention and Object Naming model. The visual attention part of the network was scaled up, from that originally used in [17], to a 32 × 32 neuron visual field Retina (input) area, with 20 × 20 neuron matrixes V1 and V2 areas and 10 × 10 neuron matrixes V4 and PFC areas for each of the 4 orientations, LIP area is a 10 × 10 neuron matrix. The auditory extension for object naming areas CB, PB, STP, and FST are modelled by 10 × 10 neuron matrixes areas, with the STP and FST consisting of the same 4 orientation-specific layers as the visual areas of the model. In our network implements 6124 simple LIF neurons, and learning is done by STDP. See Table 1 for a summary of the LIF neuron and STDP parameters. The network was implemented in the Python PyNN[1] [83] interface language for SNNs and uses the Python Brian SNN simulator[2] [84] as a backend.

4.3. Input stimuli

4.3.1. Visual input

In all experiments, visual input comes from a single iCub camera, thus avoiding the complexities of stereo processing. The images are produced in a 240 × 320 RGB format and served up via a YARP buffered image port. The Tracker module from the Aquila software architecture for cognitive robotics [82] is used to process the raw image to a saturation mask view so that objects in the scene stand out from the background and their shapes and positions can be extracted. This is further processed into an image of black and white pixels (max intensity), and resized down to the dimensions of the input layer of the visual and attentional network (32 × 32). The robot posture and visual field is fixed during learning the label and recalling phase. After activation of the LIP area in the network the robot gazes towards the salient object moving its head, neck and eyes.

The final stage of processing is to convert the white ('ON') pixels in the image to spikes for input to the visual and attentional network. There is a straight conversion of pixel to spike: the x, y pixel location in the image is mapped to a neuron ID in the network Retina input population and inserted into a spike list. This spike list is then sent to the network as a YARP Bottle object. The spikes are injected into the ON and OFF layers of the Retina area at every time step of the simulation. The same spikes are injected into both ON and OFF layers but the OFF spikes are delayed by 1.0 ms.

4.3.2. Auditory input

For the 'auditory input', the iCub robot enabled with speech-to-text and text-to-speech functionality provided by the Speech Recognition [85] and pyttsx [86] python libraries respectively running on an external PC connected to the robot's network. During learning trials the object label was repeated by the human operator and converted to text. The word stimulus is applied by sending the neuron IDs to the neural network as a YARP Bottle object, sent to a dedicated 'speech in' YARP port. The iCub interface program inserts the label into the Word Location Map linking it to the neuron IDs of the CB area. A Word Location Map linking the neuron IDs in CB and the label for the preferred object is set up at the beginning of the simulation. All other CB neurons are mapped to the label "None" in order to make it obvious if the label is not correctly associated. A group of four adjacent neurons in the CB area, of the extended auditory model for object naming, are designated to receive the 'speech' stimulus to be used for naming the preferred object (the choice of four adjacent neurons is arbitrary, it could be more or less depending upon how many words need to be encoded). Spikes were then generated and sent to the network to simulate an input spike train in the auditory network. A Poisson spike train of frequency 20 Hz is applied to these neurons at every time step of the simulation.

Artificial auditory perception is complicated, and the way in which sounds are represented in mammalian auditory cortex is not well understood [87]. A future direction of our research will be to improve and expand or auditory model, and to evaluate the possibility to use neuromorphic auditory sensors, in order to use actual auditory sound signals. For instance, the model by Coath et al. [87] learns temporal structures in auditory data and is suitable for neuromorphic implementation. Their model consist of 32 tonotopic channels interconnected through excitatory and inhibitory connections with axonal conduction delays and STDP based learning. Applications pertaining to vision, auditory and olfactory neuromorphic sensors have been discussed by [88]. Several efforts have been made to develop auditory sensors that model the human cochlea using aVLSI (analog Very Large Scale Integration). AER EAR [89] is a matched pair of silicon cochlea with an AER interface. This auditory sensor models the basilar membrane

[1] http://neuralensemble.org/PyNN/.

[2] http://briansimulator.org/.

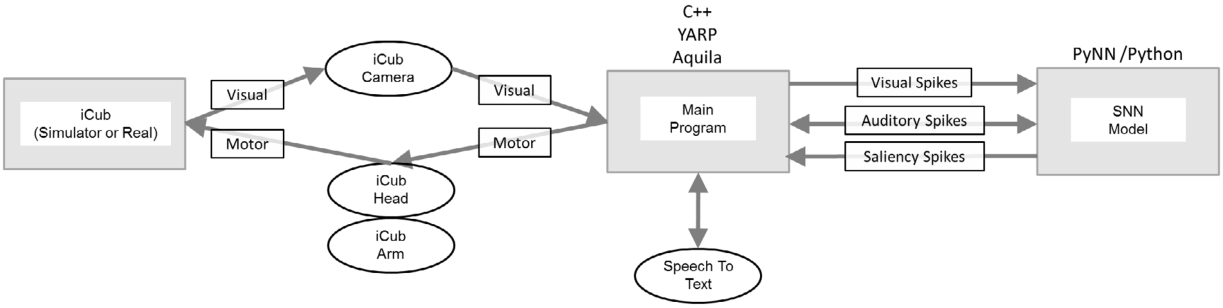
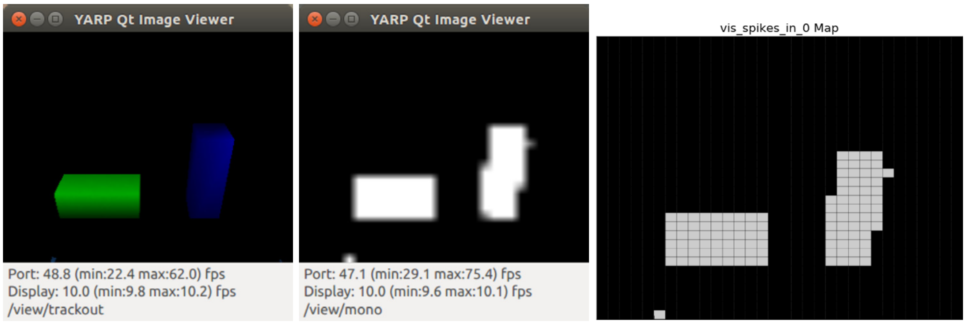


Fig. 3. Integration between the neural model and iCub, oval shapes are YARP ports, larger rectangles are software components. The 'Main' program forms the interface between the iCub and the PyNN neural network; it receives visual information and translates it to spikes as well as generating the auditory stimulus. Spikes coming back from the neural network are translated into a position so that iCub can look or point at the object. In the recall condition auditory spikes are sent back from the network and the main program maps them to a label.



Port: 48.8 (min:22.4 max:62.0) fps
Display: 10.0 (min:9.8 max:10.2) fps
/view/trackout

Port: 47.1 (min:29.1 max:75.4) fps
Display: 10.0 (min:9.6 max:10.1) fps
/view/mono

Fig. 4. Input image from iCub cameras (left). Down sampled black and white pixels (centre). Spikes input to the attentional network (right).

bio-physics by cascading low-pass filters to provide output over 32 channels [88]. AEREAR2 6658899, a further improvement of AEREAR, is 64 channel binaural audition sensor with microphone pre-amplifiers and per-channel capability has set a benchmark in neuromorphic audition [88]. Current progress will lead to developing more precise and efficient neuromorphic auditory systems by applying interesting approaches such as spike based audio front ends described in [90].

5. Experiments and results

5.1. Learning to associate a label with an object

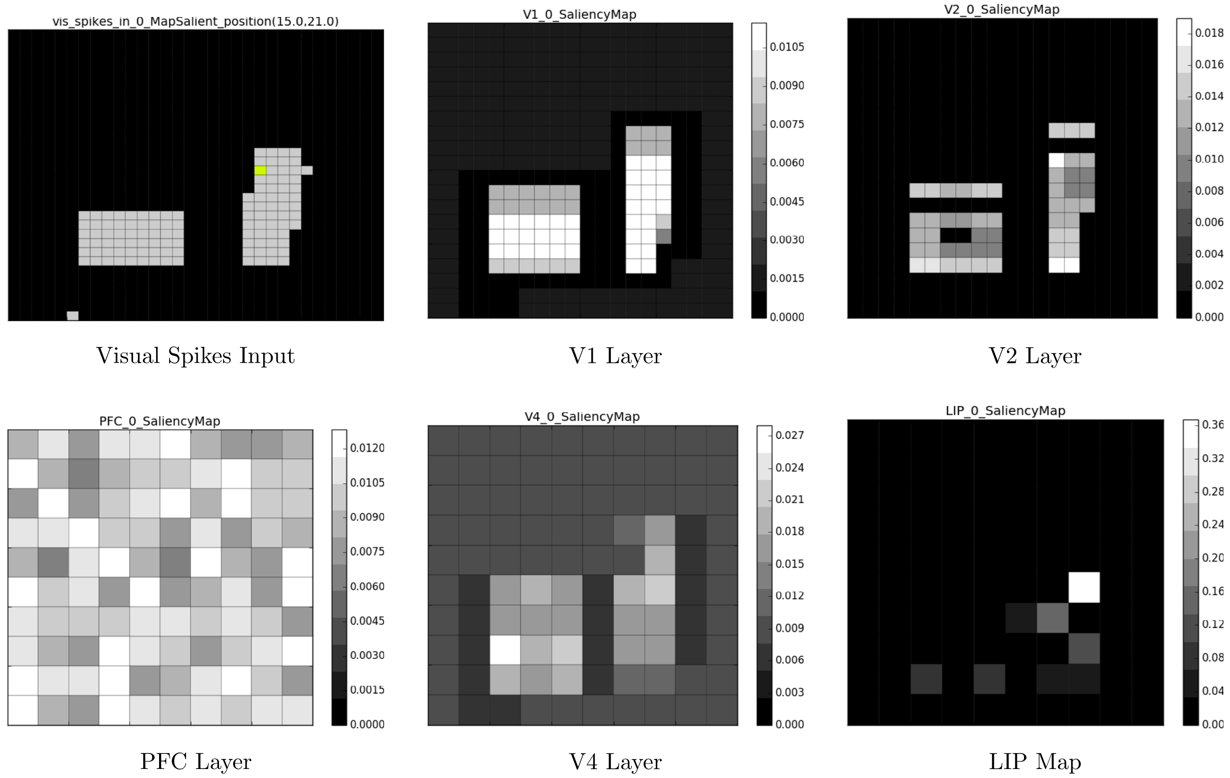
For the learning experiments, associating a 'word' label to a preferred attended object, 20 trials were run using the iCub simulator with visual and auditory stimulation for 2000 ms. For ten trials the preferred object orientation was set as Horizontal with the corresponding label "Object A" and the aversive orientation set as Vertical. For the other ten trials the preferred orientation was Vertical with the corresponding label "Object B" and the aversive orientation Horizontal. In Fig. 2 the robot setup and positioning of the objects for one trial can be seen. Fig. 4 shows the input image from the iCub simulator cameras (left). This is processed and resized into a black and white pixels image of the input layer dimensions (centre). Finally, white ('ON') pixels are converted into spikes for input to the attentional network (right).

First, we verify the operation of the visual attention part of the network by learning to attend to the preferred object, this is verified that the visual attention network produces a higher spike activity for the objects that have been biased to be 'preferred' and that the iCub robot changes its gaze so to centre the visual field on the object. Fig. 5 shows the results for learning with a 2-stimulus scene, where the vertical object is set to be the preferred object. The stimulus was provided by a pair of horizontal and vertical bars drawn in the iCub simulator. The maps show spike count, on the layer for the preferred orientation, over a run with lighter coloured areas indicating higher saliency. The V1–V2–V4 pathway selects progressively sharpened locations of visual interest. Winner-takes-all mechanism in V2 encourage the selection of a single most-salient location. The PFC provides bias for the V4 layers to prefer objects lying in one orientation but avoid objects of another orientation. The V4 layers produce a stronger input to the LIP neurons in the preferred location due to the bias effect. The LIP area winner-takes-all structure should then select a single attentional position. Spike activation can be seen in Fig. 5. Although through the network there is activity produced for both objects, the network is able to determine which one is the preferred object. The LIP saliency map shows that the vertical object was preferred and provides the more active location.

Table 2
The training procedure.

Learning to associate a label with an object
1. Set iCub in 'ready' pose (see Fig. 2)
2. Create the corresponding object in the world (random position)
3. Image of object is processed to spikes (see Fig. 4)
4. Applied 'word' stimulus to the neural network
5. Network is trained
6. Let iCub look and label object
7. Set iCub back to ready pose
8. Repeat 2–7 as required

The learning procedure, for associating a label with an object, follows as described in Table 2. First, the iCub robot is set up in its 'ready' configuration and an object is placed in the world in a random position (see Fig. 2). Visual and auditory input stimuli are provided to the neural network processed to a set of spikes. Simulation of the network is run for 2 s; the spiking activity and STDP learning throughout the visual attention network and the auditory extension leads the iCub robot to focus its view on the



Visual Spikes Input V1 Layer V2 Layer

PFC Layer V4 Layer LIP Map

Fig. 5. Visual attention learning with a 2-stimulus scene, vertical objects set to preferred. Top row left to right: The stimulus image with a pair of horizontal and vertical bars, the attended position is highlighted in the vertical object. The V1 layer for the preferred orientation. The V2 layer for the preferred orientation. Bottom row left to right: PFC is enabled biasing V4 layer. The V4 layer for the preferred orientation. The LIP saliency map, the most active location is computed by a winner-takes-all mechanism and indicates the position to attend to for the robot.

object (LIP salient point) and to learn to label the named object (STP–PB STDP enabled connection).

After each learning trial, connection weight changes between STP and PB were examined as well as activity in the FST layers. Weight differences were observed in different areas of the STP–PB weight matrix corresponding to the preferred and aversive objects. After examining connection weight changes between all STP orientation layers and PB we found that, as expected, the largest weight increases were for connections between the STP layer corresponding to the preferred direction and PB. Fig. 6 shows 2D plots of the STP–PB weight changes (ΔWeight) for connections between all orientation specific layers of STP and PB in an experiment where the vertical orientation was designated preferred. The largest weights occur on the connections for the Vertical orientation which was the preferred orientation since these neurons receive the largest activity from FST (visual stimulus from V1–V2–V4 pathway) and the PB (auditory stimulus from CB area) neurons.

Because the V4 layer corresponding to the preferred orientation receives a larger stimulus from PFC its neurons produce more spikes and this in turn causes greater weight increase on the connections from V4 to FST and between STP and PB due to STDP. The spike rate for neurons in PB that received the word stimulus increased gradually during learning, a consequence of increased connection weights between STP and PB. Fig. 7, shows a comparison spike plot of the auditory input stimulus (20 Hz) (bottom part of figure) and the spike output of the same four neurons over the course of the simulation when STDP learning is active (top part of figure). Initially the output rate is similar to that of the input stimulus but around 500 ms the rate increases until the end of the run.

Table 3 summarizes the results of running the learning experiments with 20 trials. Experiments with the Vertical objects presented greater spike activation and weight increases in the object naming part of the network than those where the Horizontal object was preferred. The average weight increase over all connections and runs was 0.4447 for the STP–PB connections and 1.4465 for the V2–V4 connections when the preferred orientation was Vertical and 0.1066 for the STP–PB connections and 1.5261 for the V2–V4 connections when the preferred orientation was set to Horizontal.

Table 3
Summarized results across all learning experiment trials.

Preferred object	STP–PB ΔWeight	PB spikes	V2–V4 ΔWeight	LIP spikes
Horizontal	0.1066	41.8	1.5261	220.9
Vertical	0.4447	58.1	1.4465	243.2

Figs. 8 and 9 show results for one learning run to label an object in the visual attention and object naming network when the preferred object to label was set to be ‘Horizontal’ and ‘Vertical’ respectively. Here the greatest activity in the LIP and FST maps are on the neurons corresponding to where the preferred object is visible in the scene; while the PB neurons activity is determined by the auditory stimulus through the CB connection. Thanks to STDP we learn the association between the auditory and visual modalities. The 2D plot of STP–PB (ΔW) shows connection weight changes between STP and PB. Largest weight increases occur in the STP–PB connections for the layer corresponding to the preferred orientation. There is no direct connection between LIP and STP, they are only indirectly associated by their connections with V4 orientation layers, thus association of the object's label is not directly dependent on location.

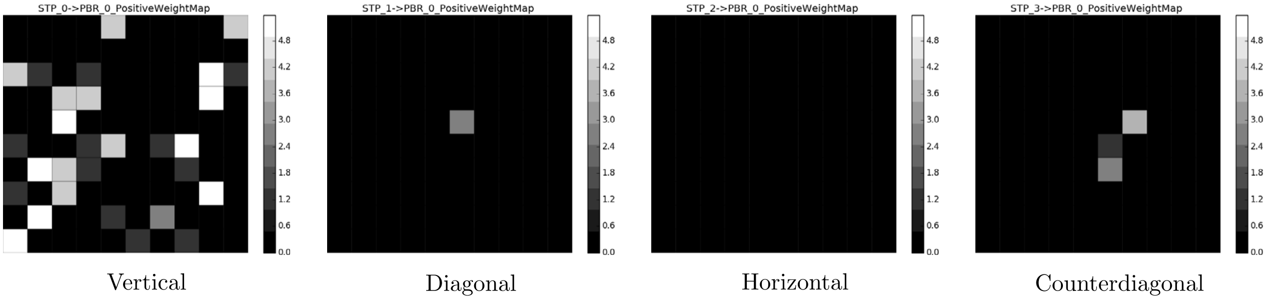


Fig. 6. Weight changes (ΔWeight $=$ final $-$ initial) for connections between all orientation specific layers of STP and PB. Left to right vertical, diagonal, horizontal and counterdiagonal orientation layers. The largest and most numerous weight increases resulted in the STP–PB connection layer for the Vertical orientation which was the preferred orientation.

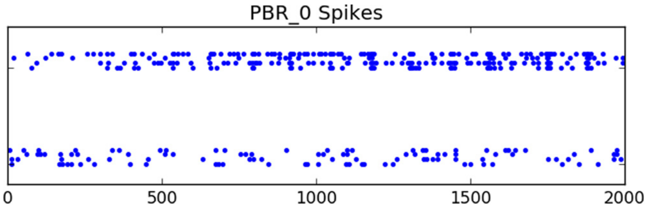


Fig. 7. Spike plots for PB area. (Bottom) PB spikes when learning is deactivated. (Top) PB spikes when learning is activated. The input 'word' stimulus for the auditory input is fixed at 20 Hz during the run. From STDP learning on the preferred connections between STP and PB, the "word neurons" received more stimulation and increased their rate when learning is active (Top) than when STDP learning is not used (Bottom).

Table 4
The testing procedure.

Recalling a previously learned name
1. Set iCub in 'ready' pose (see Fig. 2)
2. Create the corresponding objects in the world
3. Image of object is processed to spikes (see Fig. 4)
4. Disable 'word' stimulus to the neural network
5. Neural model produces a speech response
6. Active neurons IDs sent back to iCub
7. iCub translates neuron response to 'speech'
8. iCub speak object label 'word'
9. Set iCub back to ready pose
10. Repeat 2–9 as required

Recalling a previously learned name

In the recall condition, saved weights from each learning trial run in Section 5.1 were loaded into the network, STDP learning was disabled as well as the word input stimulus and only visual stimulation was applied to the network for 2000 ms. The procedure for recalling a previously learned label with an object follows as described in Table 4. First, the iCub robot is set up in its 'ready' configuration and objects are placed in the world (see Fig. 2). Visual input stimuli are provided to the neural network, processed to an appropriated set of spikes, while the auditory signal is disabled. Simulation of the network is run for 2 s; the spike activation and learned weights throughout the visual attention network and the auditory extension leads the iCub robot to focus its view on the

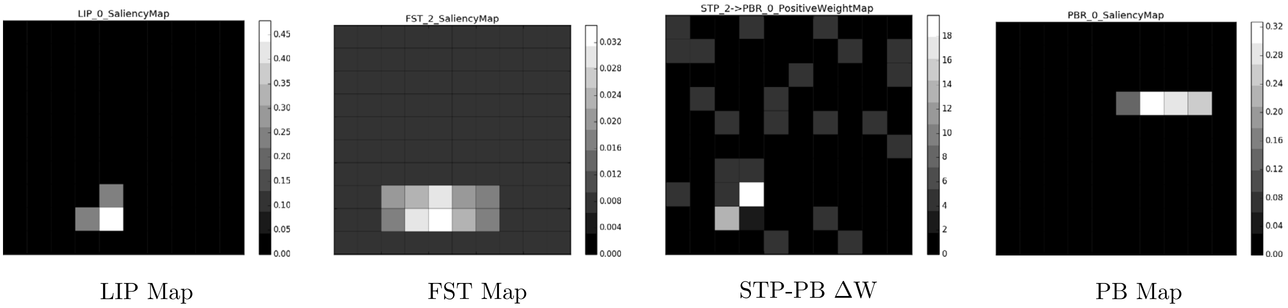


Fig. 8. Object label learning with a horizontal object set to preferred. From left to right: The LIP saliency map. The FST map for the preferred orientation layer. STP–PB connection weight changes for the preferred orientation layer. The PB saliency map.

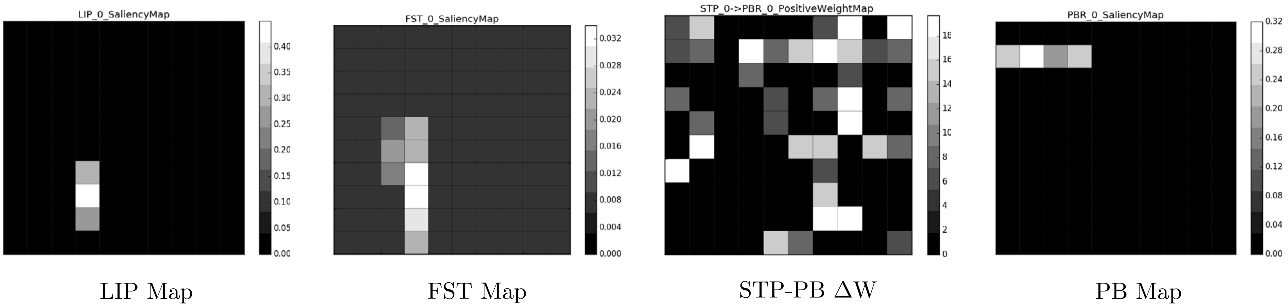


Fig. 9. Object label learning with a vertical object set to preferred. From left to right: The LIP saliency map. The FST map for the preferred orientation layer. STP–PB connection weight changes for the preferred orientation layer. The PB saliency map.

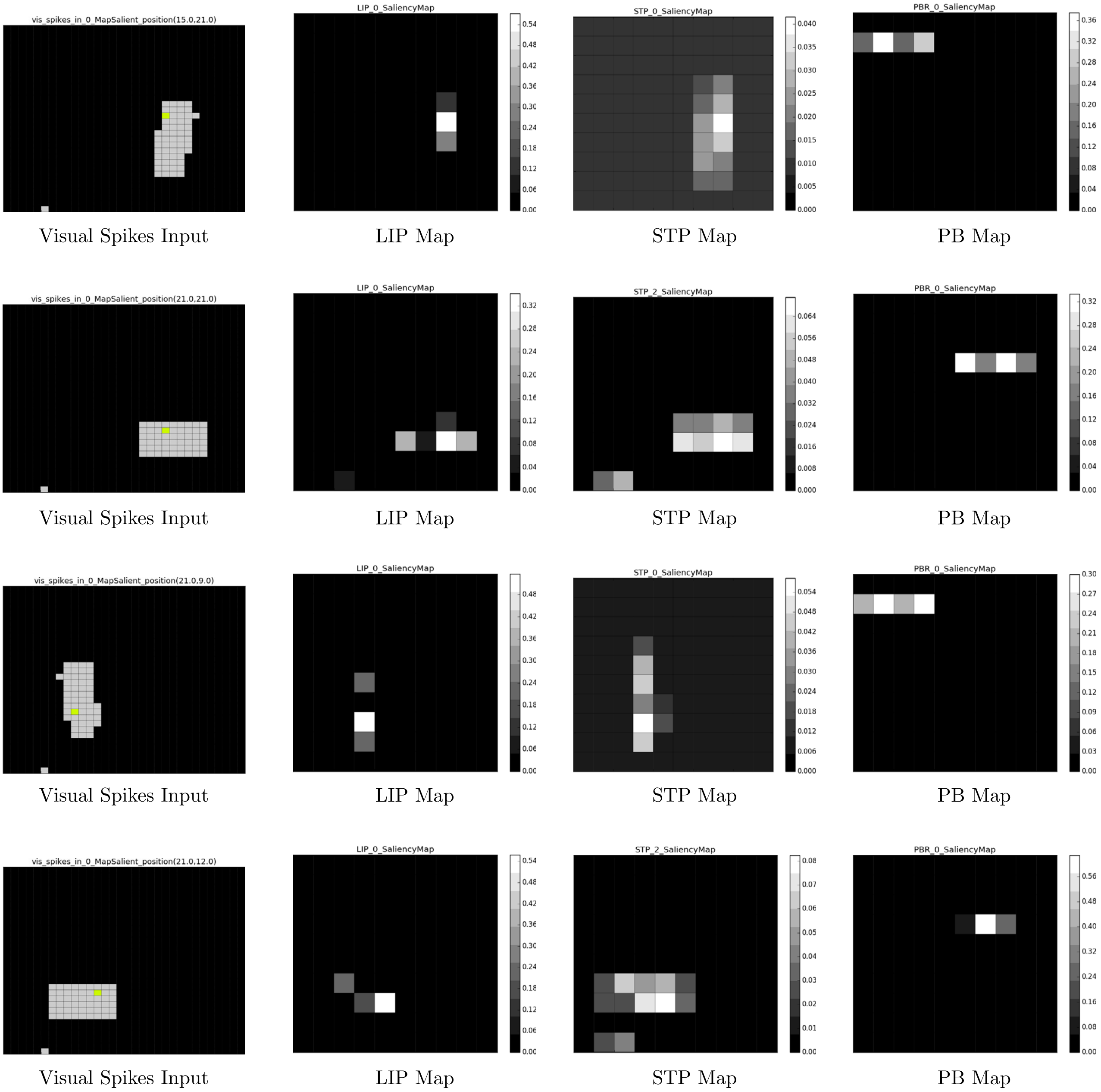


Fig. 10. Recalling a previously learned object's name. From left to right: The presented scene with the position to attend to highlighted. LIP salient point. STP saliency map. Active PB spikes of the encoded 'word' label. STP activation from the visual stimulus cause the PB neurons associated with the encoded 'word' to be active. IDs of active neurons in PB are sent back to the iCub and mapped to a label in the Word Location Map, thus recalling the object's label. (Rows 1–2) the object is on the same position that was used for learning the label. (Rows 3–4) the object was moved from its original position. In each case the label was correctly recalled as shown by the activity of the recalled PB neurons.

object (LIP salient point) and to activate FST, STP and PB neurons in the network to produce a speech response to name the object. At the end of the run the IDs of neurons that were active in PB were sent back to iCub via YARP and mapped to a label in the Word Location Map.

To verify the learning and recalling of object's name labels we performed a series of ten runs loading the saved weights from the previous learning experiments with the objects present in the scene as for the learning trials. We found that in all ten cases the correct label was recalled. On presentation of only the visual stimulus the network can still produce activation of the PB neurons needed to recall the expected object's label.

We also tested the network's ability to recall the label when the object was moved from its original position. We performed the same ten recalls but with object positions swapped and found that in all ten cases the correct label was recalled despite the object being in a different location from where it was originally named. Fig. 10 shows examples of positions of the object during recall.

To further test recall when the object was moved from its original position recall was done with the object moved to several different positions and also including another object (with different orientation) in the scene as a distractor. Fig. 11 shows examples of positions of the object during recall. In all cases the label was correctly recalled when the object was attended to.

5.3. Real speech IO on the iCub robot

The experiments described in Sections 5.1 and 5.2 were repeated on the iCub robot enabled with speech-to-text and text-to-speech functionality running on a laptop connected to the robot's

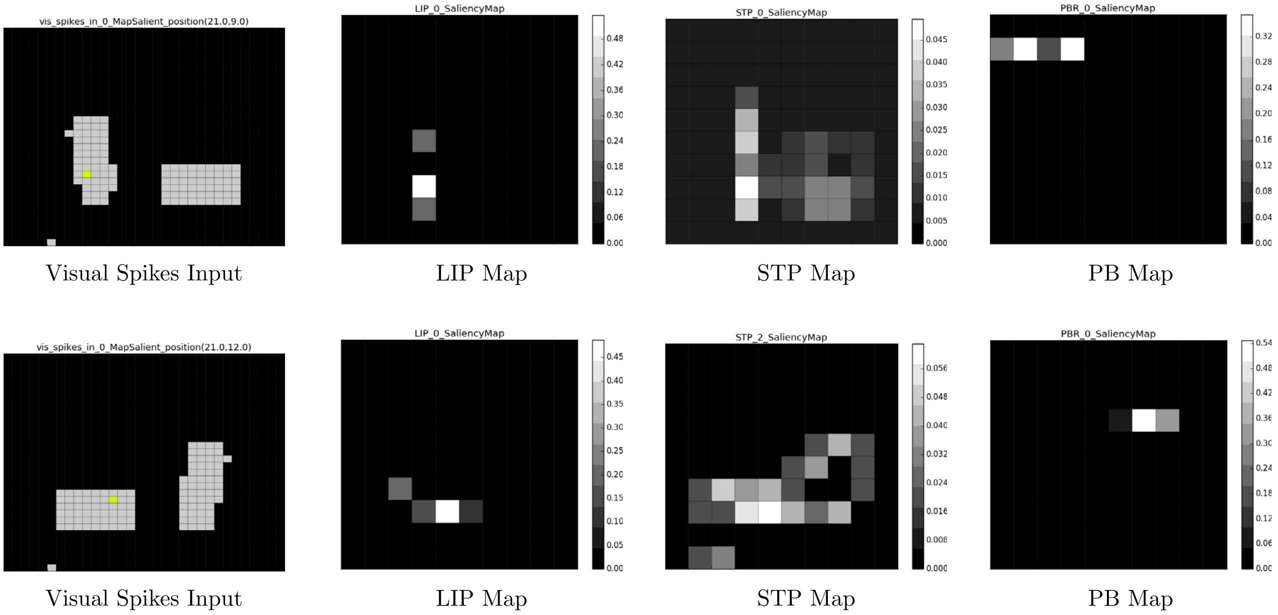
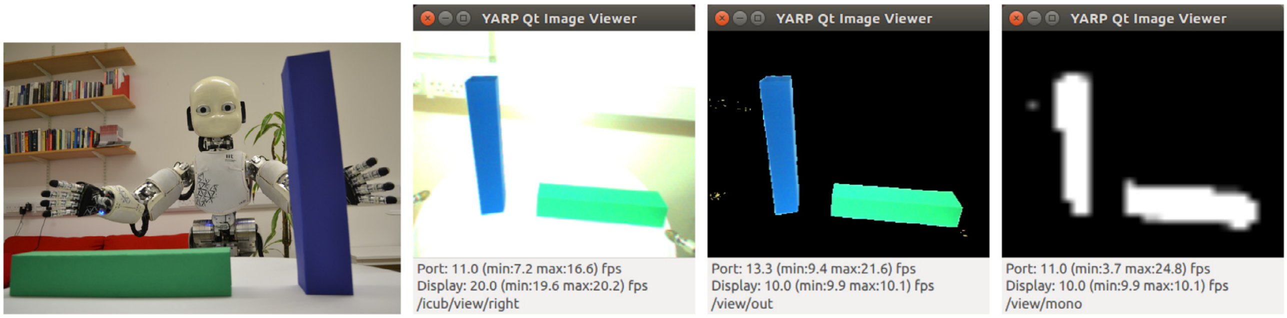


Fig. 11. Recalling a previously learned object's name when an additional object is included in the scene as a distractor. Top-row a run of the experiment where the preferred orientation was set to 'Vertical'. Bottom-row a run of the experiment where the preferred orientation was set to 'Horizontal'.



Port: 11.0 (min:7.2 max:16.6) fps
Display: 20.0 (min:19.6 max:20.2) fps
/icub/view/right

Port: 13.3 (min:9.4 max:21.6) fps
Display: 10.0 (min:9.9 max:10.1) fps
/view/out

Port: 11.0 (min:3.7 max:24.8) fps
Display: 10.0 (min:9.9 max:10.1) fps
/view/mono

Fig. 12. Real iCub Humanoid Robot Setup. Raw input image from iCub cameras. Saturation mask processing for objects shape and position extraction. Down sampled black and white pixels.

Table 5
Summarized results of learning experiment trials with iCub robot.

Preferred object	STP–PB ⊿Weight	PB spikes	V2–V4 ⊿Weight	LIP spikes
Horizontal	0.1049	47.8	1.4088	173.4
Vertical	0.4073	61.8	1.4491	206.8

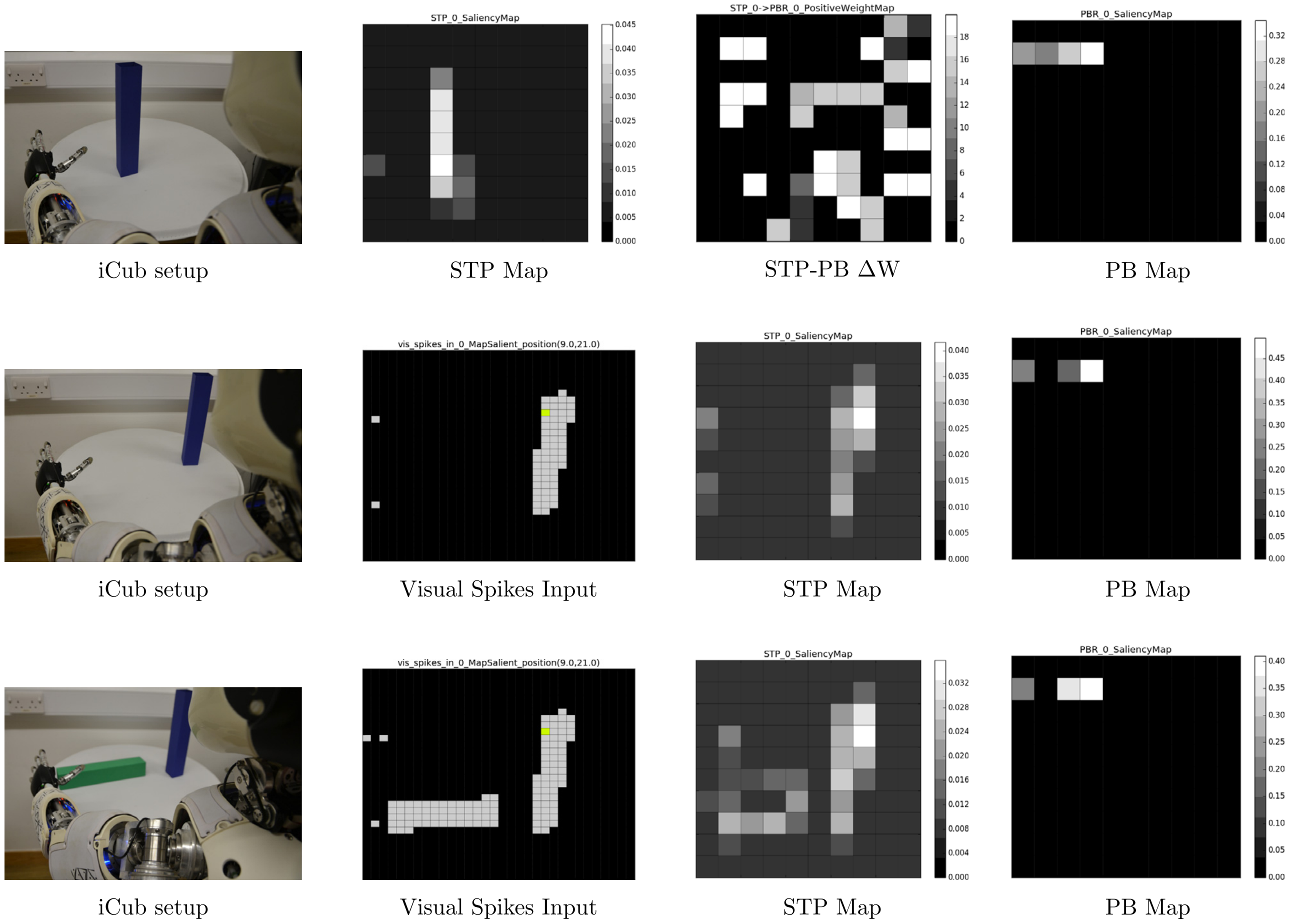
local network. During learning trials the object label was repeated by the human operator, converted to text and sent to a dedicated 'speech in' YARP port. The iCub interface program inserted the label into the Word Location Map linking it to the neuron IDs of 4 CB neurons. Spikes were then generated and sent to the network. For recall, active neuron IDs were sent back to iCub and mapped to the label stored in the Word Location Map as before. This text was sent to a dedicated YARP 'speech out' port and converted to actual speech. Fig. 12 shows the real iCub set up for the experiments and the processing of the input image with the real iCub cameras for transformation into spikes for input to the attentional network.

As before ten trials were done where one run consisted of a learning phase followed by a recall phase. The biasing of 'preferred'/'non-preferred' stimuli were the same as described before. The results were similar to those produced with the iCub simulator in that in all cases weight increases through V2–V4 and STP–PB connections were greater for the preferred orientation.

The average weight increase over all connections and runs was 0.4073 for the STP–PB connections and 1.4088 for the V2–V4 connections when the preferred orientation was Vertical and 0.1049 for the STP–PB connections and 1.4088 for the V2–V4 connections when the preferred orientation was set to Horizontal. Table 5 summarizes the results of running the experiments with the iCub robot.

For the recall phase of the experiments, in all cases the label was recalled correctly when the preferred object was attended to. To further test recall when the object was moved from its original position a learning run with a single (Vertical) object in the scene was done and then recall was done with the object moved to several different positions and also including another object (Horizontal) in the scene as a distractor. Figs. 13 and 14 show experimental runs for both Vertical and Horizontal preferred objects. The figure shows the positions of the object when learning the object's label and also recall examples where the position of the object was changed and also when an additional object was present in the scene as a distractor. In all cases the label was correctly recalled when the object was attended to.

In the recall phase STP activation from the visual stimulus caused the PB neurons associated with the encoded 'word' to be active. IDs of active neurons in PB were sent back to the iCub and mapped to a label in the Word Location Map and converted into speech.



iCub setup STP Map STP-PB ΔW PB Map

iCub setup Visual Spikes Input STP Map PB Map

iCub setup Visual Spikes Input STP Map PB Map

Fig. 13. Word learning and recall in different object positions for 'Vertical' preferred objects. The learning phase occurred with the object as in the top row. Successful recall of the word and attention to the object occurred when the object was moved to a different location and also when another object was added to the scene.

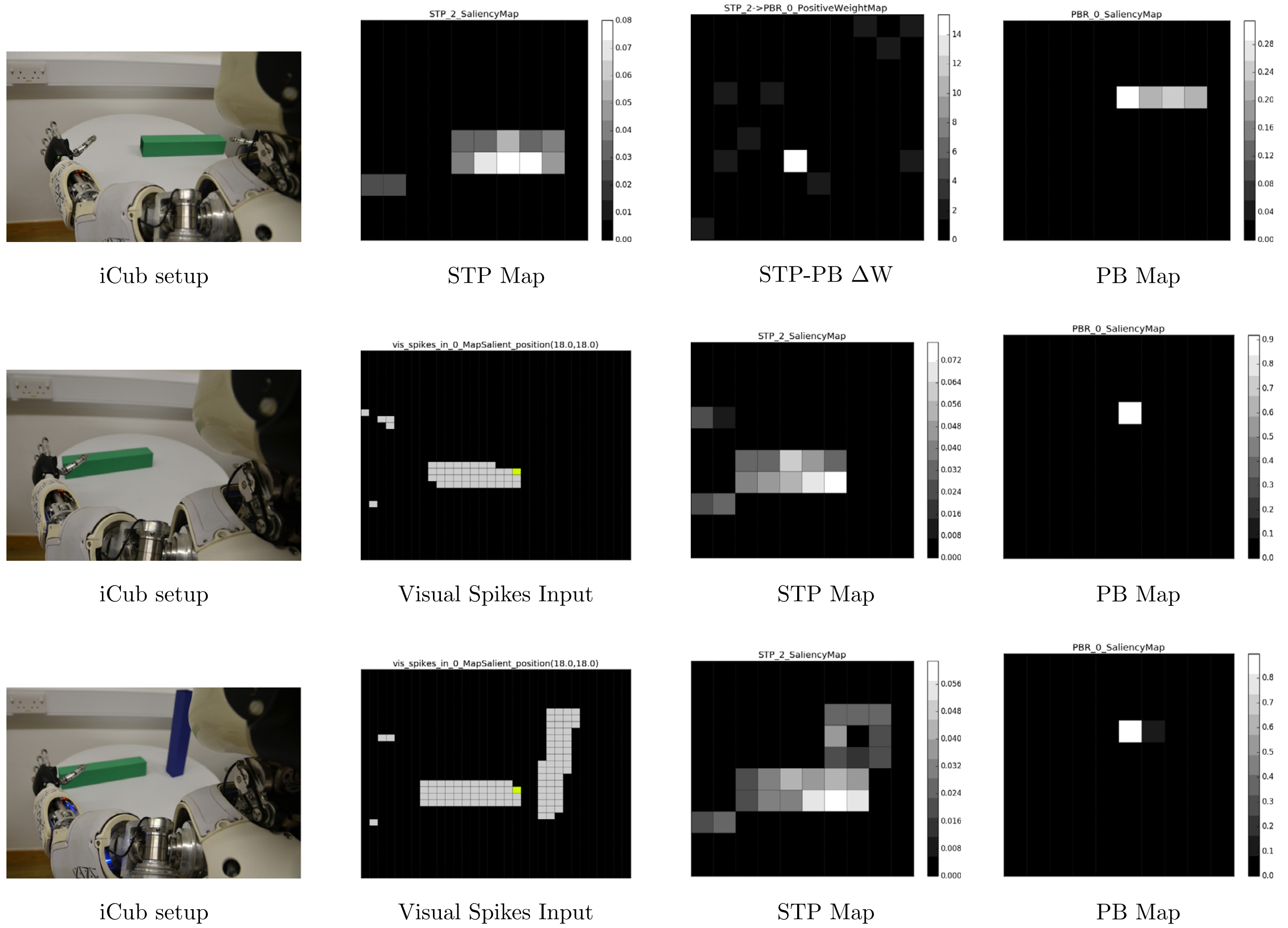
6. Discussion and conclusions

In the work described here a neuroanatomically grounded object naming system based upon a visual attention network was developed. It demonstrates that a label could be associated with a salient object via Spike-Timing Dependent Plasticity in a simple system. This combined auditory and visual attention system forms a basis from which to add other modalities, for example motor areas, along the lines of the neurobiological model of Garagnani et al. [6].

The visual attention model architecture of Adams et al. [17] was extended to model real brain areas of the auditory cortex to incorporate the capability of naming the attended objects using Spike-Timing Dependent Plasticity. The extended auditory modality associates a word with an object of preferred orientation. Extra subareas, with plausible neuroanatomy of multisensory and auditory neurons, were added to the attention network to link visual and auditory information. Two multi-modal and two auditory areas were added to the network. We model the core–belt–parabelt auditory processing pathway with two areas, CB and PB. The combined auditory area (CB) models the primary auditory cortex and auditory belt, while the PB area represents the Parabelt, and forms the link to the multi-modal areas, subdivision and connections of auditory system are supported by [58,59]. The multi-modal areas represent the Superior Temporal Sulcus which is subdivided in the model into two areas, the Fundus of the Superior Temporal Sulcus (FST) and the Superior Temporal Polysensory (STP) area. The subdivision and connections of STS multi-modal area into caudal (FST) and rostral (STP) areas are supported by [59,60]. V4 visual area connections to the FST are supported in [61] by anatomical evidence that such a connection exists in presence primates.

For learning, 4 neurons in CB corresponding to a 'word' are stimulated with Poisson spikes (20 Hz) at the same time as the visual input is processed. We showed how connection weights between the V4 area (representing the object's orientation) and the FST and STP areas (multi-modal system) increased substantially more for the orientations biased as preferred, and that learning of the salient object's label occurs between STP and the PB area (auditory cortex) thanks to STDP. After learning, when word input is disabled, the activation from V4 through FST–STP–PB causes word activation to occur, thus once learning has taken place visual stimulation alone could recall a label associated with the object. We have also shown that recall occurred even when the object was moved to a different position, hence 'word' association is learned positionally independent.

This work has described a neurorobotics approach based upon neuroanatomically grounded Spiking Neural Networks with biologically-inspired learning for object naming driven by visual attention. It provides a proof-of-concept case for the integration of biologically inspired neural networks with robotics for basic language acquisition, as visual attention is crucial for learning object names. Most important for robotics, the model shows pattern completion ability; after training if just the visual input is presented the auditory parts of the pattern are completed successfully. Our experiments point the way to one of the goals of cognitive robotics: self-directed robots able to respond adaptively and appropriately rather than imperatively to the combination of unexpected events and indeterminate consequences characteristic of the real world.



iCub setup STP Map STP-PB ΔW PB Map

iCub setup Visual Spikes Input STP Map PB Map

iCub setup Visual Spikes Input STP Map PB Map

Fig. 14. Word learning and recall in different object positions for 'Horizontal' preferred objects. The learning phase occurred with the object as in the top row. Successful recall of the word and attention to the object occurred when the object was moved to a different location and also when another object was added to the scene.

We intend to add several enhancements in future work. Following on from our previous work in [17] the network will be implemented on SpiNNaker hardware enabling us to scale up the network whilst maintaining as close to real time speed as possible. Using a larger network (2 to 4 times the number of neurons used now up to tens of thousands of neurons) will make it possible to have more objects in the scene as distractors and also more complex objects. Also, we plan to implement the network using event-driven cameras and neuromorphic sensors. Applications pertaining to vision, auditory and olfactory neuromorphic sensors have been discussed by [88]. Key contributions such as DVS and DAVIS cameras and AEREAR AEREAR2 cochleas have provided considerable progress towards a sensor design that simulates neurobiological vision and auditory sensing. A limitation of the current network is that objects can only be recognized in terms of their orientation, so it will be necessary to add extensions to enable preference by colour as well as orientation and a more complex shape description. It is important to be able to use a greater repertoire of objects with richer visual features so that more interesting learning experiments can be done. In the experiments described here, PFC orientation bias is hardcoded in the system, however, it is already possible to learn orientation preferences rather than hardcode them and so a similar mechanism could be implemented for orientation and colour combined. The enhanced system will be used to model the developmental pathway for language, moving from single word to two word associations — for example, action-verb and noun.

Acknowledgements

This work was supported by EPSRC (the UK Engineering and Physical Sciences Research Council) under grants EP/J004561/1 and EP/J00457X/1 (BABEL).

References

[1] M. Garagnani, F. Pulvermüller, Neuronal correlates of decisions to speak and act: Spontaneous emergence and dynamic topographies in a computational model of frontal and temporal areas, Brain Lang. 127 (1) (2013) 75–85. http://dx.doi.org/10.1016/j.bandl.2013.02.001.

[2] G. Pezzulo, L. Barsalou, A. Cangelosi, M. Fischer, K. McRae, M. Spivey, Computational grounded cognition: a new alliance between grounded cognition and computational modeling, Front. Psychol. 3 (2013) 612. http://dx.doi.org/10.3389/fpsyg.2012.00612.

[3] F. Pulvermüller, M. Garagnani, T. Wennekers, Thinking in circuits: toward neurobiological explanation in cognitive neuroscience, Biol. Cybernet. 108 (5) (2014) 573–593. http://dx.doi.org/10.1007/s00422-014-0603-9.

[4] J. Zhong, A. Cangelosi, S. Wermter, Toward a self-organizing pre-symbolic neural model representing sensorimotor primitives, Front. Behav. Neurosci. 8 (2014) 22. http://dx.doi.org/10.3389/fnbeh.2014.00022.

[5] M. Garagnani, F. Pulvermüller, Conceptual grounding of language in action and perception: a neurocomputational model of the emergence of category specificity and semantic hubs, Eur. J. Neurosci. 43 (6) (2016) 721–737.

[6] M. Garagnani, G. Lucchese, R. Tomasello, T. Wennekers, F. Pulvermüller, A spiking neurocomputational model of high-frequency oscillatory brain responses to words and pseudowords, Front. Comput. Neurosci. 10 (2017) 145. http://dx.doi.org/10.3389/fncom.2016.00145.

[7] F. Pulvermüller, L. Fadiga, Active perception: Sensorimotor circuits as a cortical basis for language, Nat. Rev. Neurosci. 11 (5) (2010) 351–360.

[8] L.W. Barsalou, Grounded cognition, Annu. Rev. Psychol. 59 (1) (2008) 617–645.

[9] F. Pulvermüller, Brain mechanisms linking language and action, Nat. Rev. Neurosci. 6 (2005) 576–582.

[10] F. Pulvermüller, Words in the brain's language, Behav. Brain Sci. 22 (2) (1999) 253–279.

[11] M.A. Arbib, Mirror system activity for action and language is embedded in the integration of dorsal and ventral pathways, Brain Lang. 112 (1) (2010) 12–24.

[12] P.F. Dominey, T. Inui, Cortico-striatal function in sentence comprehension: Insights from neurophysiology and modeling, Cortex 45 (8) (2009) 1012–1018.

[13] M. Garagnani, T. Wennekers, F. Pulvermüller, A neuroanatomically grounded Hebbian-learning model of attention-language interactions in the human brain, Eur. J. Neurosci. 27 (2) (2008) 492–513.

[14] A.F. Morse, T. Belpaeme, A. Cangelosi, L.B. Smith, Thinking with your body: modelling spatial biases in categorization using a real humanoid robot, in: Proceedings Cognitive Science Conference, 2010, pp. 1362–1368.

[15] D. Caligiore, A.M. Borghi, D. Parisi, R. Ellis, A. Cangelosi, G. Baldassarre, How affordances associated with a distractor object affect compatibility effects: A study with the computational model tropicals, Psychol. Res. 77 (1) (2013) 7–19.

[16] S.V. Adams, T. Wennekers, A. Cangelosi, M. Garagnani, F. Pulvermüller, Learning visual-motor cell assemblies for the icub robot using a neuroanatomically grounded neural network, in: 2014 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain, CCMB, 2014.

[17] S.V. Adams, A.D. Rast, C. Patterson, F. Galluppi, K. Brohan, J.-A. Pérez-Carrasco, T. Wennekers, S. Furber, A. Cangelosi, Towards real-world neurorobotics: Integrated neuromorphic visual attention, in: Neural Information Processing: 21st International Conference, ICONIP 2014, 2014, pp. 563–570.

[18] C. Yu, L.B. Smith, Embodied attention and word learning by toddlers, Cognition 125 (2) (2012) 244–262. http://dx.doi.org/10.1016/j.cognition.2012.06.016.

[19] D.O. Hebb, The Organization of Behavior: A Neuropsychological Theory, Wiley, New York, 1949.

[20] S. Song, K.D. Miller, L.F. Abbott, Competitive Hebbian learning through spike-timing-dependent synaptic plasticity, Nature Neurosci. 3 (2000) 919–926.

[21] A.F. Morse, V.L. Benitez, T. Belpaeme, A. Cangelosi, L.B. Smith, Posture affects how robots and infants map words to objects, PLoS One 10 (3) (2015) 1–17. http://dx.doi.org/10.1371/journal.pone.0116012.

[22] F. Galluppi, K. Brohan, S. Davidson, T. Serrano-Gotarredona, J.-A.P. Carrasco, B. Linares-Barranco, S. Furber, A real-time, event-driven neuromorphic system for goal-directed attentional selection, in: Neural Information Processing: 19th International Conference, ICONIP 2012, Doha, Qatar, November 12–15, 2012, Proceedings, Part II, 2012, pp. 226–233.

[23] F. Kaplan, Neurorobotics: an experimental science of embodiment, Front. Neurosci. 2 (2008) 23. http://dx.doi.org/10.3389/neuro.01.023.2008.

[24] J.L. Krichmar, J. Conradt, M. Asada, Neurobiologically inspired robotics: Enhanced autonomy through neuromorphic cognition, Neural Netw. 72 (2015) 1–2. http://dx.doi.org/10.1016/j.neunet.2015.11.004.

[25] J. Krichmar, H. Wagatsuma, Neuromorphic and Brain-Based Robots, Cambridge University Press, New York, NY, USA, 2011.

[26] M. Rucci, D. Bullock, F. Santini, Integrating robotics and neuroscience: brains for robots, bodies for brains, Adv. Robot. 21 (10) (2007) 1115–1129.

[27] B. Cox, J. Krichmar, Neuromodulation as a robot controller, IEEE Robot. Autom. Mag. (2009) 72–80.

[28] R. de Azambuja, A. Cangelosi, S.V. Adams, Diverse, noisy and parallel: a new spiking neural network approach for humanoid robot control, in: International Joint Conference on Neural Networks, IJCNN, 2016, pp. 1134–1142.

[29] D. Gamez, A.K. Fidjeland, E. Lazdins, iSpike: a spiking neural interface for the iCub robot, Bioinspir. Biomim. 7 (2) (2012) 008–025.

[30] P. Barros, D. Jirak, C. Weber, S. Wermter, Multimodal emotional state recognition using sequence-dependent deep hierarchical features, Neural Netw. 72 (2015) 140–151.

[31] G. Park, J. Tani, Development of compositional and contextual communicable congruence in robots by using dynamic neural network models, Neural Netw. 72 (2015) 109–122. http://dx.doi.org/10.1016/j.neunet.2015.09.004.

[32] K. Seepanomwan, D. Caligiore, A. Cangelosi, G. Baldassarre, Generalisation, decision making, and embodiment effects in mental rotation: A neurorobotic architecture tested with a humanoid robot, Neural Netw. 72 (2015) 31–47. http://dx.doi.org/10.1016/j.neunet.2015.09.010.

[33] M. Beyeler, N. Oros, N. Dutt, J.L. Krichmar, A GPU-accelerated cortical neural network model for visually guided robot navigation, Neural Netw. 72 (2015) 75–87. http://dx.doi.org/10.1016/j.neunet.2015.09.005.

[34] F. Walter, F. Röhrbein, A. Knoll, Neuromorphic implementations of neurobiological learning algorithms for spiking neural networks, Neural Netw. 72 (2015) 152–167. http://dx.doi.org/10.1016/j.neunet.2015.07.004.

[35] N. Kasabov, E. Capecci, Spiking neural network methodology for modelling, classification and understanding of EEG spatio-temporal data measuring cognitive processes, Inform. Sci. 294 (2015) 565–575. http://dx.doi.org/10.1016/j.ins.2014.06.028.

[36] E. Smith, M.S. Lewicki, Efficient coding of time-relative structure using spikes, Neural Comput. 17 (1) (2005) 19–45. http://dx.doi.org/10.1162/0899766052530839.

[37] Y. Kuniyoshi, L. Berthouze, Neural learning of embodied interaction dynamics, Neural Netw. 11 (78) (1998) 1259–1276. http://dx.doi.org/10.1016/S0893-6080(98)00085-9.

[38] M.G. Gaskell, M. Hare, W.D. Marslen-Wilson, A connectionist model of phonological representation in speech perception, Cogn. Sci. 19 (4) (1995) 407–439. http://dx.doi.org/10.1207/s15516709cog1904_1.

[39] M.F. Joanisse, M.S. Seidenberg, Impairments in verb morphology after brain injury: A connectionist model, Proc. Natl. Acad. Sci. USA 96 (13) (1999) 7592–7597.

[40] D.C. Plaut, L.M. Gonnerman, Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing?, Lang. Cogn. Process. 15 (4–5) (2000) 445–485.

[41] M.H. Christiansen, N. Chater, Connectionist psycholinguistics: capturing the empirical data, Trends Cogn. Sci. 5 (2) (2001) 82–88. http://dx.doi.org/10.1016/S1364-6613(00)01600-4.

[42] G. Westermann, E.R. Miranda, A new model of sensorimotor coupling in the development of speech, Brain Lang. 89 (2) (2004) 393–400. http://dx.doi.org/10.1016/S0093-934X(03)00345-6.

[43] F.H. Guenther, S.S. Ghosh, J.A. Tourville, Neural modeling and imaging of the cortical interactions underlying syllable production, Brain Lang. 96 (3) (2006) 280–301. http://dx.doi.org/10.1016/j.bandl.2005.06.001.

[44] K. Brohan, K. Gurney, P. Dudek, Using reinforcement learning to guide the development of self-organised feature maps for visual orienting, in: K. Diamantaras, W. Duch, L.S. Iliadis (Eds.), Artificial Neural Networks –ICANN 2010: 20th International Conference, Thessaloniki, Greece, September 15-18, 2010, Proceedings, Part II, Springer, Berlin, Heidelberg, 2010, pp. 180–189. http://dx.doi.org/10.1007/978-3-642-15822-3_23.

[45] F. Ponulak, A. Kasinski, Introduction to spiking neural networks: Information processing, learning and applications, Acta Neurobiol. Exp. 71 (4) (2011) 409–433.

[46] W. Gerstner, W. Kistler, Spiking Neuron Models: An Introduction, Cambridge University Press, New York, NY, USA, 2002.

[47] W. Maass, Networks of spiking neurons: The third generation of neural network models, Neural Netw. 10 (9) (1997) 1659–1671. http://dx.doi.org/10.1016/S0893-6080(97)00011-7.

[48] H. Paugam-Moisy, S. Bohte, Computing with spiking neuron networks, in: G. Rozenberg, T. Bäck, J.N. Kok (Eds.), Handbook of Natural Computing, Springer, Berlin, Heidelberg, 2012, pp. 335–376. http://dx.doi.org/10.1007/978-3-540-92910-9_10.

[49] A.D. Rast, F. Galluppi, X. Jin, S.B. Furber, The leaky integrate-and-fire neuron: A platform for synaptic model exploration on the SpiNNaker chip, in: The 2010 International Joint Conference on Neural Networks, IJCNN, 2010, pp. 1–8. http://dx.doi.org/10.1109/IJCNN.2010.5596364.

[50] L.F. Abbott, S.B. Nelson, Synaptic plasticity: taming the beast, Nature Neurosci. 3 (2000) 1178–1183.

[51] S. Song, L. Abbott, Cortical development and remapping through spike timing-dependent plasticity, Neuron 32 (2) (2001) 339–350. http://dx.doi.org/10.1016/S0896-6273(01)00451-2.

[52] D. Hubel, Eye, Brain, and Vision, in: Scientific American Library Series, Henry Holt and Company, 1995.

[53] M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex, Nature Neurosci. 2 (1999) 1019–1025.

[54] M.A. Lebedev, A. Messinger, J.D. Kralik, S.P. Wise, Representation of attended versus remembered locations in prefrontal cortex, PLoS Biol. 2 (11) (2004). http://dx.doi.org/10.1371/journal.pbio.0020365.

[55] B. Noudoost, M.H. Chang, N.A. Steinmetz, T. Moore, Top-down control of visual attention, Curr. Opin. Neurobiol. 20 (2) (2010) 183–190. http://dx.doi.org/10.1016/j.conb.2010.02.003.

[56] M.L. Platt, P.W. Glimcher, Neural correlates of decision variables in parietal cortex, Nature 400 (6741) (1999) 233–238. http://dx.doi.org/10.1038/22268.

[57] U. Pattacini, Modular cartesian controllers for humanoid robots: Design and implementation on the icub, Dizertacná Práca, Istituto Italiano Di Tecnologia, 2011.

[58] J.H. Kaas, T.A. Hackett, M.J. Tramo, Auditory processing in primate cerebral cortex, Curr. Opin. Neurobiol. 9 (2) (1999) 164–170. http://dx.doi.org/10.1016/S0959-4388(99)80022-1.

[59] J.H. Kaas, T.A. Hackett, Subdivisions of auditory cortex and processing streams in primates, Proc. Natl. Acad. Sci. 97 (22) (2000) 11793–11799. http://dx.doi.org/10.1073/pnas.97.22.11793.

[60] H.-O. Karnath, New insights into the functions of the superior temporal cortex, Nature Rev. Neurosci. 2 (8) (2001) 568–576. http://dx.doi.org/10.1038/35086057.

[61] L.G. Ungerleider, T.W. Galkin, R. Desimone, R. Gattass, Cortical connections of area v4 in the macaque, Cerebral Cortex 18 (3) (2008) 477–499.

[62] L. Steels, The symbol grounding problem has been solved. So what's next?, in: M. de Vega (Ed.), Symbols and Embodiment: Debates on Meaning and Cognition, Oxford University Press, Oxford, 2008.

[63] A. Cangelosi, M. Schlesinger, Developmental Robotics: From Babies to Robots, The MIT Press, 2014.

[64] J. Saunders, C.L. Nehaniv, C. Lyon, The acquisition of word semantics by a humanoid robot via interaction with a human tutor, in: K. Dautenhahn, J. Saunders (Eds.), New Frontiers in Human-Robot Interaction, John Benjamins Publishing Company, 2011, pp. 211–234.

[65] J. Saunders, C.L. Nehaniv, C. Lyon, Robot learning of lexical semantics from sensorimotor interaction and the unrestricted speech of human tutors, in: Second International Symposium on New Frontiers in Human-Robot Interaction, AISB Convention, Leicester, UK, 2010.

[66] J. Saunders, H. Lehmann, F. Foerster, C. Nehaniv, Robot acquisition of lexical meaning: Moving towards the two-word stage, in: 2012 IEEE International Conference on Development and Learning and Epigenetic Robotics, ICDL, 2012. http://dx.doi/org/10.1109/DevLrn.2012.6400588.

[67] C. Lyon, C.L. Nehaniv, J. Saunders, Preparing to talk: interaction between a linguistically enabled agent and a human teacher, in: AAAI Fall Symposium, 2010.

[68] C. Lyon, C.L. Nehaniv, J. Saunders, Interactive language learning by robots: The transition from babbling to word forms, PLoS One 7 (6) (2012) e38236. http://dx.doi.org/10.1371/journal.pone.0038236. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3374830/.

[69] F. Foerster, J. Saunders, C. Nehaniv, Robots that say 'no'. Affective symbol grounding and the case of intent interpretations, IEEE Trans. Cogn. Dev. Syst. (2017).

[70] L. Steels, Evolving grounded communication for robots, Trends Cogn. Sci. 7 (7) (2003) 308–312. http://dx.doi.org/10.1016/S1364-6613(03)00129-3. URL http://www.sciencedirect.com/science/article/pii/S1364661303001293.

[71] L. Steels, Language games for autonomous robots, IEEE Intell. Syst. 16 (5) (2001) 16–22. http://dx.doi.org/10.1109/5254.956077.

[72] L. Steels, T. Belpaeme, coordinating perceptually grounded categories through language: a case study for colour, Behav. Brain Sci. 28 (4) (2005) 469489. http://dx.doi.org/10.1017/S0140525X05000087.

[73] T. Nakamura, T. Araki, T. Nagai, N. Iwahashi, Grounding of word meanings in latent dirichlet allocation-based multimodal concepts, Adv. Robot. 25 (17) (2011) 2189–2206. http://dx.doi.org/10.1163/016918611X595035.

[74] M. Attamimi, Y. Ando, T. Nakamura, T. Nagai, D. Mochihashi, I. Kobayashi, H. Asoh, Learning word meanings and grammar for verbalization of daily life activities using multilayered multimodal latent Dirichlet allocation and Bayesian hidden Markov models, Adv. Robot. 30 (11–12) (2016) 806–824. http://dx.doi.org/10.1080/01691864.2016.1172507.

[75] T. Araki, T. Nakamura, T. Nagai, S. Nagasaka, T. Taniguchi, N. Iwahashi, Online learning of concepts and words using multimodal LDA and hierarchical Pitman-Yor language model, in: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012, pp. 1623–1630. http://dx.doi.org/10.1109/IROS.2012.6385812.

[76] A. Cangelosi, Grounding language in action and perception: From cognitive agents to humanoid robots, Phys. Life Rev. 7 (2) (2010) 139–151. http://dx.doi.org/10.1016/j.plrev.2010.02.001. URL http://www.sciencedirect.com/science/article/pii/S1571064510000187.

[77] F. Broz, C.L. Nehaniv, T. Belpaeme, A. Bisio, K. Dautenhahn, L. Fadiga, T. Ferrauto, K. Fischer, F. Frster, O. Gigliotta, S. Griffiths, H. Lehmann, K.S. Lohan, C. Lyon, D. Marocco, G. Massera, G. Metta, V. Mohan, A. Morse, S. Nolfi, F. Nori, M. Peniak, K. Pitsch, K.J. Rohlfing, G. Sagerer, Y. Sato, J. Saunders, L. Schillingmann, A. Sciutti, V. Tikhanoff, B. Wrede, A. Zeschel, A. Cangelosi, The italk project: a developmental robotics approach to the study of individual, social, and linguistic learning, Top. Cogn. Sci. 6 (3) (2014) 534–544. http://dx.doi.org/10.1111/tops.12099.

[78] C. Lyon, C.L. Nehaniv, J. Saunders, T. Belpaeme, A. Bisio, K. Fischer, F. Frster, H. Lehmann, G. Metta, V. Mohan, A. Morse, S. Nolfi, F. Nori, K. Rohlfing, A. Sciutti, J. Tani, E. Tuci, B. Wrede, A. Zeschel, A. Cangelosi, Embodied language learning and cognitive bootstrapping: Methods and Design Principles, Int. J. Adv. Robot. Syst. 13 (3) (2016) 105. http://dx.doi.org/10.5772/63462.

[79] S. Coradeschi, A. Loutfi, B. Wrede, A short review of symbol grounding in robotic and intelligent systems, KI-Künstliche Intelligenz 27 (2) (2013) 129–136.

[80] M.A. Arbib, G. Metta, P. van der Smagt, Neurorobotics: From Vision to Action, in: B. Siciliano, O. Khatib (Eds.), Springer Handbook of Robotics, Springer, Berlin, Heidelberg, 2008, pp. 1453–1480. http://dx.doi.org/10.1007/978-3-540-30301-5_63.

[81] S.B. Furber, D.R. Lester, L.A. Plana, J.D. Garside, E. Painkras, S. Temple, A.D. Brown, Overview of the spinnaker system architecture, IEEE Trans. Comput. 62 (12) (2013) 2454–2467. http://dx.doi.org/10.1109/TC.2012.142.

[82] M. Peniak, A. Morse, A. Cangelosi, Aquila 2.0 software architecture for cognitive robotics, in: 2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics, ICDL, 2013, pp 1–6. http://dx.oi.org/d10.1109/DevLrn.2013.6652565.

[83] A.P. Davison, D. Brüderle, J. Eppler, J. Kremkow, E. Muller, D. Pecevski, L. Perrinet, P. Yger, PyNN: a common interface for neuronal network simulators, Front. Neuroinform. 2 (2008).

[84] D. Goodman, R. Brette, The brian simulator, Front. Neurosci. 3 (2009) 26. http://dx.doi.org/10.3389/neuro.01.026.2009. URL https://www.frontiersin.org/article/10.3389/neuro.01.026.2009.

[85] A. Zhang, Speech recognition, URL https://github.com/Uber/speech_recognition#readme.

[86] RapidWareTech, pyttsx. URL https://github.com/RapidWareTech/pyttsx.

[87] M. Coath, S. Sheik, E. Chicca, G. Indiveri, S. Denham, T. Wennekers, A robust sound perception model suitable for neuromorphic implementation, Front. Neurosci. 7 (2014) 278. http://dx.doi.org/10.3389/fnins.2013.00278.

[88] A. Vanarse, A. Osseiran, A. Rassau, A review of current neuromorphic approaches for vision, auditory, and olfactory sensors, Front. Neurosci. 10 (2016) 115. http://dx.doi.org/10.3389/fnins.2016.00115.

[89] V. Chan, S.C. Liu, A. van Schaik, Aer ear: A matched silicon cochlea pair with address event representation interface, IEEE Trans. Circuits Syst. I. Regul. Pap. 54 (1) (2007) 48–59. http://dx.doi.org/10.1109/TCSI.2006.887979.

[90] T.J. Koickal, R. Latif, L. Gouveia, E. Mastropaolo, S. Wang, A. Hamilton, R. Cheung, M. Newton, L. Smith, Design of a spike event coded RGT microphone for neuromorphic auditory systems, in: 2011 IEEE International Symposium of Circuits and Systems, ISCAS, 2011, pp. 2465–2468. http://dx.doi.org/10.1109/ISCAS.2011.5938103.

Daniel Hernández García is a postdoctoral research fellow at the Centre for Robotics and Neural Systems at the University of Plymouth. He received his M.S. degree in Robotics and Automation and Ph.D. degree in Electrical Engineering, Electronics and Automation from the Universidad Carlos III of Madrid, in 2010 and 2014, respectively. His research interests include Cognitive and Neuro-Robotics, Artificial Intelligence, Human–Robot Interaction, Machine Learning Algorithms and Robot Perception and Language Learning.

Samantha Adams received the B.Sc. degree in Mathematics and Physics from the Open University, UK, in 2003 and the MRes degree (with Distinction) in Computing from the University of Plymouth, UK in 2009. She gained a Ph.D. in Computational Neuroscience from the University of Plymouth in 2013. She has worked as a scientific software engineer for many years in various domains and recently as a Postdoctoral Researcher affiliated to Plymouth University. Her research interests focus on biologically inspired computing, encompassing machine learning and AI and how techniques from these fields can be applied to make smarter applications.

Alexander Rast is a Senior Research Fellow with the department of Electronics and Computer Science at the University of Southampton. His current work examines innovative methods for configuring large parallel systems without shared memory. Previously he worked at the APT group at the University of Manchester investigating associative language learning on the iCub humanoid robot with biologically realistic spiking neural network models implemented on the SpiNNaker chip. He received his Ph.D. in Computer Science from the University of Manchester in 2011. Before going to the University of Manchester, he was Research Director for Inficom, Inc, a Seattle, USA-based startup company developing advanced systems for wireless communications using neural control. His research interests include protocols for neuromorphic communications and formal design and implementation methods in real-world applications for problems that can be mapped as graph topologies including spiking neural networks.

Thomas Wennekers studied Physics at the Heinrich-Heine University (Duesseldorf, Germany) and Computer Science at the University of Ulm (Germany), where he received a Ph.D. in Computer Science in 1997. He was postdoctoral research fellow at the Max Planck Institute for Mathematics in the Sciences (Leipzig) from 1999 to 2003, and Juniorprofessor in Theoretical Neuroscience at the Ruhr-University Bochum. Since November 2003 he is Reader in Computational Neuroscience at Plymouth University (UK). His research interests are large-scale spiking neuron models of sensory, perceptual and cognitive functions, and their application in future computing technologies.

Steve Furber CBE FRS FREng is ICL Professor of Computer Engineering at the University of Manchester, UK. After completing a B.A. in mathematics and a Ph.D. in aerodynamics at the University of Cambridge, UK, he spent the 1980s at Acorn Computers, where he was a principal designer of the BBC Microcomputer and the ARM 32-bit microprocessor. Over 90 billion variants of the ARM processor have since been manufactured, powering much of the world's mobile and embedded computing. At Manchester since 1990, he leads the SpiNNaker project, which is delivering a computer incorporating a million ARM processors for brain modelling applications.

Angelo Cangelosi is Professor of Artificial Intelligence and Cognition and Director of the Centre for Robotics and Neural Systems at the University of Plymouth. Cangelosi received a master degree (Laurea) in Psychology at University of Rome La Sapienza, and a Ph.D. in psychology and Artificial Intelligence, which led him to specialize in AI and robotics. Cangelosi's research expertise is on cognitive developmental robotics, human–robot interaction and artificial intelligence. He has produced over 250 publications, edited four books, and is first author of the 2015 book "Developmental Robotics" (MIT Press). He is currently the coordinator of the H2020 Marie Curie EID "APRIL" and PI in two other Marie Curie networks (DCOMM, SECURE). He has coordinated two FP7 projects (ITALK IP and RobotDoc ITN) and two large UK projects (VALUE and BABEL). In 2012–13 was Chair of the IEEE CIS Technical Committee on Autonomous Mental Development.