Natural human-robot interaction through spatial language: a Dynamic Neural Field approach

Yulia Sandamirskaya, John Lipinski, Ioannis Iossifidis and Gregor Schöner
Institut für Neuroinformatik, Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail: yulia.sandamirskaya|ioannis.iossifidis|gregor.schoner@ini.rub.de, 2johnlipinski@gmail.com

***Abstract*— For an autonomous robotic system, the ability to share the same workspace and interact with humans is the basis for cooperative behavior. In this work, we investigate human spatial language as the communicative channel between the robot and the human, facilitating their joint work on a tabletop. We specifically combine the theory of Dynamic Neural Fields that represent perceptual and cognitive states with motor control and linguistic input in a robotic demonstration. We show that such a neural dynamic framework can integrate across symbolic, perceptual, and motor processes to generate task-specific spatial communication in real time.**

I. Introduction

One goal of human-robot interaction is to enable joint action in an environment shared between non-expert humans and autonomous robots. Those environments are unstructured, partially unknown, and include the human operator, with whom the robotic assistant must interact and cooperate. In this case, two fundamental requirements arise. First, the robotic assistant must possess a certain degree of autonomy, obtaining information about the environment through its perceptual processes and acting in relation to objects in the environment. Second, the robotic assistant must be guided by the human operator in real time and with an intuitive user interface. Such guiding requires a two-way interaction, with human instructions being picked up by the robot autonomously and, conversely, feedback about the robot's state, or requests for additional information being directed to the user.

The traditional means of providing inputs to an electronic device – with a keyboard, a computer mouse, or touch panels – are not practical in human-robot interaction. The need for natural and intuitive interfaces is widely recognized (see, for instance, [1], [2]). A number of labs have developed robotic assistants that interact with a human user by analyzing linguistic commands and gestures [3], [4], [5], [6]. Many of the successful approaches emphasize the need for a common representation of both linguistic and perceptual information. Tight integration of language processing, decision making, learning, perception, and sensory-motor control is desirable for performant human-robot systems, but has not been achieved to the same extent. Here we propose a theoretical framework that is inspired by analogies with the nervous system and that enables the systematic integration across these functions.

The described aspects of human-robot interaction can be probed in a scenario where a robot and a human user share a tabletop workspace. The robot assists the human in, for instance, assembly, cooking, or cleaning-up. The robot might hand over an object, which is out of reach for the human, or perform sub-tasks with the human supervision.

On a tabletop workspace, the notion of space is essential. It is not only the basis for object localization, it can also be used to support object selection in ambiguous situations. Spatial language is the natural means to exchange information about spatial relations ("left", "right", "behind") with a human cooperator. It is remarkable, how flexible and powerful human spatial language is [7], [8]. Indeed, a task as simple as specifying the location of, say, an apple relative to a laptop involves several cognitive processes: detecting the presence of objects in a scene, selecting a reference object relative to which the spatial term is defined, aligning the spatial reference frame with this object, and making the decision about either the identity or location of the target object. Moreover, these processes are guided by sensory information, which may be distorted by eye- and head-movements, new objects appearing in view, occlusions, or simply sensory noise.

The use of human spatial language in human-robot interaction is constrained by the properties of the underlying cognitive processes. Dynamic Field Theory (DFT) [9] is a neural-dynamic framework, in which cognitive models can be formulated in an embodied and dynamical way, enabling their coupling to real robotic sensors and motors. Recently, we introduced a DFT architecture to account for flexible human spatial language behaviors. The DFT spatial language model was first designed to account for metric biases, established in empirical research on human spatial language [10]. We then further developed this model in a robotic implementation to demonstrate how the empirical model can be connected to a real-world sensory information and generate flexibly real time spatial-language behaviors [11].

Here, we extensively test the DFT spatial language architecture in a human-robot interaction scenario in which human and robot share a tabletop workspace. The robot is able to answer questions about the colors of the objects on the tabletop and spatial relations between them. The user can also direct the robot's attention to a particular object by specifying it's color or spatial relation and the color of a reference object, relative to which the relation is to be applied. In ambiguous situations, the user might want to specify both color and spatial relation of an object, or let the robot decide autonomously which object to chose from the ones that satisfy the provided constraint. The initial cue and the disambiguating cue are provided by the human user in a natural way, no special timing is needed. In particular, a corrective signal can be given when the robot makes an erroneous decision based on the ambiguous initial information. In this work, we extend the motor repertoire of the robot with a pointing gesture, in order to provide better feedback to the user about the robot's decisions on the object of interest, and to probe the capability of the dynamic field representations to guide robotic movement.

The neural-dynamic model is implemented on an anthropomorphic robotic platform [25]. Perceptual input from a robotic camera shapes an attractor of a neurally inspired dynamic field that represents the visual scene. Categorical linguistic input about the identities of objects, or about the spatial relations between them is integrated in the graded neural fields' dynamics. These dynamics stabilize decisions about the presence of objects, their features, or spatial relations, but stay sensitive to relevant perceptual and linguistic input. This results in a flexible and fluent communication with the user, in which the dynamics of the architecture adapts naturally to the user input and the dynamical visual scene. Moreover, the graded metric DFT representation of the visual scene can guide the robotic action by setting an attractor for the dynamics controlling the position of an actuator.

The paper is organized as follows. In the remainder of the Introduction, we review the related work. The second section provides a brief introduction to DFT and a description of the DFT spatial language architecture. In the third section, we present the robotic platform CoRA, on which two demonstrations were carried out. The demonstrations exemplify two paths in user-robot interaction about objects on the tabletop, emphasizing performance of the model in ambiguous situations. We conclude with a short discussion of what is achieved in this work.

A. *Related Work*

Enabling human-robot interaction and making it more natural for the non-expert human user is currently a topic of vivid research. The robot companion introduced in [12] detects the human operator and picks-up his or her commands. The robot combines the information provided by speech and pointing. A key issue in that work is to coordinate multiple input modalities. Orchestrating the algorithms that support speech recognition, sound source localization, and image processing is the main bottleneck for a flexible and robust communication between the user and the robot.

A similarly complex architecture [6] relies on human-robot interaction to enable learning of visual features and associating them with language categories such as color or shape names and spatial relations. Extensive human supervision is needed, in particular, when the categories must be adjusted in changing environments.

A combined Bayesian and symbolic architecture enables a robotic assistant to learn word-to-meaning associations from interaction with a human user based on the robot's representations of its actions [5]. Because these representations are thus associated with words, they can be used to instruct the robot to perform specific tasks. These representations also provided context for speech recognition.

Roy and colleagues [13], [14] emphasize the internal representations that link spatial semantics to visual processing and motor behaviors, highlighting role of embodiment and situatedness. Ripley, the most recent variant of this approach [4], achieves an impressive range of language behaviors in a tabletop setting similar to our scenario. Our concepts overlap with those employed in these systems, but use analogue neuronal dynamics rather than algorithms throughout.

Another aspect of spatial language is the perspective alignment between two interacting robots [15], as well as the emergence of spatial referencing within a population [16].

II. METHODOLOGICAL BACKGROUND

In this section, we briefly review the framework of Dynamic Field Theory (DFT) and the spatial language architecture built in this framework.

A. Dynamic Fields Theory (DFT)

The neurally-based representational language of DFT [17] emphasizes attractor states and their instabilities in the dynamics of neural fields. According to DFT, cognitive states may be defined in terms of distributions of activation that evolve over continuous dimensions. These dimensions represent parameters of the state (e.g. color, position, angular deviation, speed). The activation function characterizes presence of the particular parameter values at every moment in time.

1) The dynamic field equation: The activation within dynamic neural fields evolves in time according to a non-linear differential equation (1) that defines the rate of change, $\dot{u}(\mathbf{x},t)$, of the activation function, $u(\mathbf{x},t)$, defined over a metric dimension, $\mathbf{x}$, at each moment in time.

$$\begin{aligned}\tau\dot{u}(\mathbf{x},t) = & - u(\mathbf{x},t) + h + I(\mathbf{x},t) \\ & + \int f\Big(u(\mathbf{x}',t)\Big)\omega(\Delta\mathbf{x})d\mathbf{x}'\end{aligned} \quad (1)$$

According to this equation, the activity in a dynamic neural field converges with a time-constant, τ, to an attractor, defined by a negative resting level, h, external inputs, $I(\mathbf{x},t)$, and lateral interactions within the field.

The lateral interactions are homogeneous within the field, they are thus expressed by a convolution term. The convolution kernel is a Gaussian function:

$$\omega(\Delta\mathbf{x}) = c_{exc}e^{\frac{-(\Delta x)^2}{2\sigma^2}} - c_{inh}, \quad \Delta x = x - x'. \quad (2)$$

This kernel means that every two near-by sites of a neural field, x and x', are connected with a positive (excitatory) connection weight, and distant sites are connected with a negative (inhibitory) weight. The short-range excitation is of amplitude c_{exc} and range σ. The global inhibition is of amplitude c_{inh}.

The sigmoid non-linearity, Equation (3), shapes the output of the dynamic field. The sites of the field with positive outputs are termed to be active.

$$f(u) = \frac{1}{1+e^{-\beta u}} \quad (3)$$

2) Specificity of the field's dynamics: Because of the negative resting level, the field is quiescent (not active) without external input. When the external input exceeds an activation threshold, a localized bump of activity evolves in the field [18]. These activity bumps represent features, locations, action plans in DFT.

The lateral interactions in the field determine the stability of the localized bump solution. The lateral interactions also provide for the instabilities in the field's dynamics, which result in detection, memory, selection among alternatives, or "forgetting" of dynamic attractor states [9].

3) Applications: Recent applications of DFT include empirical research in visual working memory [19], the development of spatial working memory [20], and spatial semantic processing [10]. In a robotic context, DFT has been employed for object recognition [21], cooperation [22], sequence generation [23].

Its successful application in both empirical and robotic contexts suggests that DFT provides a suitable framework for realizing the representational integration supporting spatial language semantics in human-robot interaction.

B. *The DFT spatial language framework*

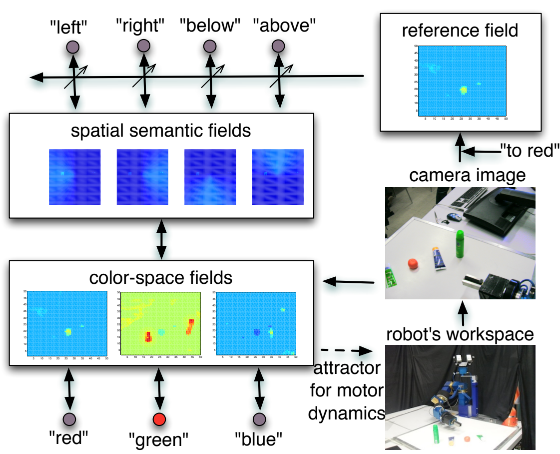


Fig. 1. Overview of the architecture.

This section outlines the overall structure and dynamics of the DFT spatial language architecture (Figure 1). Along with dynamic neural fields, we also introduce here discrete dynamical neural nodes that represent the linguistic input provided by the human user, as well as the linguistic output of the robot.

The robotic camera provides input to a set of dynamic neural fields that represent the visual space through associations of colors and their locations. Each of these *color-space fields* (see Figure 1) receives input from pixels in the camera image, the hue value of which falls within a certain range, corresponding to a basic color (e.g. blue, red, green, etc). The resolution of color is low here because only a few colors are needed to represent the objects. Thus, several discrete two-dimensional color-space fields are sufficient to sample the visual scene. The continuum of colors can be resolved to an arbitrary degree of granularity in a three-dimensional implementation of the color-space field.

The visual input from the camera alone is not sufficient to activate the color-space fields. Language input specifying the color of the object boosts the resting level of the corresponding color-space field ("green" on Figure 1). When summed with the visual input from the camera, this activation boost induces an instability that leads to formation of a single peak of activation centered over the object that provides the strongest input of the specified color. The spatial language input also influences the color-space fields' dynamics through the spatial semantic fields (see below).

A peak of activation in a color-space field sets an attractor for the dynamics that controls the robotic movements (see section III).

The *reference field* (Figure 1) is a spatially tuned dynamic field, which also receives visual input. When the user specifies the reference object's color, the corresponding "reference-color" neural node becomes active and pixels with the specified color in the camera image provide input to the reference field. A reference field activation peak specifies the location of the reference object and continuously tracks its position. It also filters out irrelevant inputs and camera noise, thus stabilizing the reference object representation. Having a stable, but nonetheless updatable reference object representation allows the spatial semantics to be continuously aligned with the visual scene.

The spatial terms are characterized by the shape of *spatial semantic templates*, which define the connectivity between a particular spatial term and a "retinotopic" space. These connection weights are modulated by the localized activity in the reference field representing the user-defined reference object. Thus, the spatial semantic templates are aligned with the location of the reference object in the image. This shift is accomplished by a convolution of the outcome of the reference field with the spatial semantic template. The particular functions defining "left", "right", "in front", and "behind" here are two-dimensional Gaussians in polar coordinates and are based on a neurally-inspired approach to English spatial semantic representation [24]. In Cartesian coordinates, they have a tear-drop shape. The shapes of spatial semantic templates were explicitly designed in this work, but arbitrary shapes can be acquired in a learning process, by laying down memory traces in the neural fields' dynamics.

The *spatial semantic fields* (see Figure 1) are dynamic neural fields with weak lateral interactions. They integrate the spatial semantic user input (aligned with the reference object) with the summed output of the color-space fields. Both these connections are reciprocal. The summed output of the spatial semantic fields serves as input to the color-space fields, enhancing activation in those regions corresponding to the specific spatial term. When active, they also provide input to spatial-term nodes, triggering a spatial-term output ("left", "right", etc.).

III. Robotic implementation

The model described in the previous section was implemented on the anthropomorphic robot CoRA [25]. Experiments demonstrate the successful generation of pointing and head movements consistent with the constraints provided by the user, as well as production of correct answers to questions about colors of objects and spatial relations between objects in a scene.

A. CoRA

The robotic assistant system CoRA has an anthropomorphic seven degrees of freedom (DoF) arm mounted on a one DoF trunk. CoRA is built as a modular robotic system, in which each module is servo-controlled and communicates via a CAN-bus interface with the controlling PC. Above the trunk, a two DoF pan/tilt unit carrying a stereo color camera system is assembled (Figure 1).

B. CoRA head movement

The movement of the head of the robot is governed by an attractor dynamics defined over pan and tilt angles. An activity peak in the color-space fields is projected on the horizontal and the vertical axis of the image plane. The resulting activity profiles are then multiplied with a linear monotonically increasing function. These products yield larger values for activity peaks farther away from the midline of the visual field. These values are then used as the target forcelets in the dynamics of the head's pan and tilt. These forcelets effectively set an attractor at the head pose, which centers the target object in the camera view. The dynamics for pan and tilt are run sequentially. First, the pan converges to center the target along the horizontal axes of the image plane. Only then tilt is changed. This form of servo control eliminates the need for the transformation from image-based to robot-based coordinates.

C. CoRA arm movement

The dynamics controlling the head movement is also used to control the end-effector position of the robotic arm. The pointing gesture is performed in a plane parallel to the tabletop. Assuming a fixed distance to the tabletop, we use the tilt angle of the robot head to calculate the distance from the robot trunk to the pointing position. The pan angle of the head is used to calculate the direction to the target location in the horizontal plane. Combining both, we calculate the desired Cartesian position of the end-effector for the pointing gesture. The inverse kinematics problem for the redundant robot arm is solved in closed form [26].

IV. Demonstrations

To illustrate how this system works and to probe the dynamic processes supporting representational integration in spatial semantics, we describe here two particular demonstrations. In both demonstrations, four objects are put on the tabletop (Figure 2): a green tube of cream, a red plastic apple, a blue tube of sunscreen, and a green deo-spray can. The objects are represented by their colors in the current implementation. Thus, the two green objects cannot be distinguished by the robot, and additional information from the user is needed to act in the task. The color modality is a placeholder in this implementation of a more complete object recognition system [21]. However, the failure to disambiguate two objects would require support from the user even when such system is used.

In Demonstration 1, the user asks the robot to point to the green object. The user additionally specifies the spatial relation "to the left from the red one", when the robot chooses the (wrong) green object to the right from the red one, which is slightly more salient. In Demonstration 2, the user asks the robot to point to the object to the right from the red object, adding "the green one", when the robot first selects the more salient blue object, which also satisfies the specified spatial relation to the red object. The color of the reference object (red) is provided by the user in both demonstrations.

In the current implementation, the user provides inputs about colors of objects and spatial relations through a graphical user interface, whereas this interface can be easily replaced by a keywords based speech recognition system. An important peculiarity of the interface is that both linguistic and graded inputs and outputs are possible: the user can either provide a language term for a color, or set the particular hue on a color-wheel, the same is true for the robot. For the purposes of this paper, however, the most important aspect of human-robot interaction is the flexibility with which the user can provide inputs when needed, or desired. These inputs are integrated into the architecture's dynamics in real time and cause meaningful effects by shifting the attractor states of the dynamics.

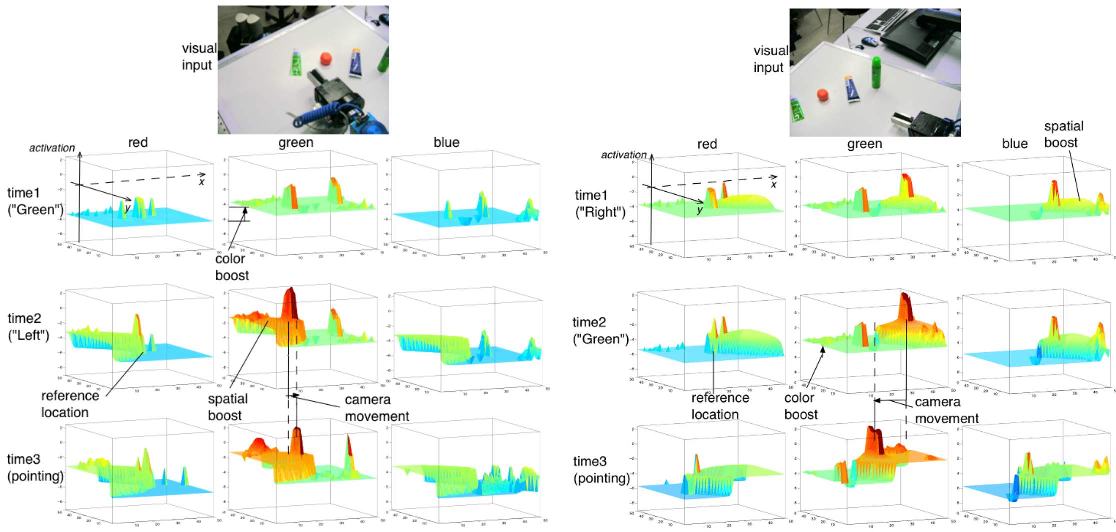
A. Demonstration 1: Specifying "green" then "left"

The user input in Demonstration 1 is analogous to saying "Point at the green object, the one to the left of the red object". When "green" is first specified as the color of the target object, the "green" color-term node activation provides a homogeneous activation boost to the "green" color-space field (see Figure 2a, upper row). Because there are two green objects in the camera image, the "green" color-space field reaches activation threshold at both the locations of the green tube and the green can. Although the lateral inhibition in the dynamical field slows down the formation of an activation peak in this field, the green can provides stronger input and finally wins the competition. As a consequence, a peak evolves at that location in the color-space field and the camera starts to move to center this object. A peak also evolves in the "right" spatial semantic field, and the robot signals selection of the object to the right from the red object.

When the user then specifies "left" as the spatial relation, however, the "left" spatial semantic field receives a homogeneous boost, providing additional activation to the left portion of the color-space fields (Figure 2a, middle row). This additional input provides an advantage to the representation of the green tube in the "green" color-space field. Because of the lateral inhibition in the field, the current peak at the (incorrect) location of the green can is eventually extinguished and a peak at the correct location emerges. The new peak in the green field resets the attractor for the dynamics of camera and arm movements. The robot turns the camera to center the correct green object in the image (Figure 2a, lower row), and points to it.

B. Demonstration 2: Specifying "right" then "green"

The user input in Demonstration 2 is analogous to saying "Point at the object to the right of the red object, the green one". When the user first specifies "right" as the spatial relation, the positive activation in the "right" spatial semantic field is propagated to the color-space fields. Here, that region overlaps with the location of the blue tube and the green can (Figure 2b, upper row). The overlap with the blue tube is slightly stronger because it lies closer to the reference object. A peak of activation builds-up in the blue color-space field, the robot signals about



(a) Demonstration 1

(b) Demonstration 2

Fig. 2. Snapshots of activity of the color-space dynamic fields at decisive time-points in the dynamics during the two robotic demonstrations (see text for details).

selection of the blue object and initiates the head and arm movements.

When the user defines "green" as the color of the object of interest (Figure 2b, middle row), however, the uniform activation boost to the "green" color-space field provides advantage to this fields and a localized activation peak builds up at the green can location. The inhibitory interaction between the color-space fields forces the peak at the (wrong) blue item location to extinguish. The activity in the "green" color-space field resets the attractor for the dynamics of motor control for both camera and arm movements. The camera centers the correct green object in the camera image and the robot starts the pointing movement towards it (Figure 2b, lower row).

V. Discussion

The presented work is the first step towards a language system in robots that is grounded in the neural dynamics of a cognitive spatial language architecture. The DFT spatial language model was proven to capture several traits of human spatial language processing [10], [11], and work continues to address additional empirical findings. Our demonstrations reveal how neural-dynamic theories such as DFT can successfully integrate differing representational dimensions to generate autonomous spatial language behaviors in robots.

The robotic system in our demonstrations was able to answer questions related to the object's position, color, its spatial relation to other objects in the workspace, and to point at such objects in order to confirm the decision. The user could provide information about the object of interest at any time during the interaction, flexibly shaping the dynamics of the architecture, emphasizing either the feature, or spatial aspects of the visual scene. Interaction with the robot is thus more natural for the human user, who does not have to provide all the necessary information before the task execution can start.

Clearly, a truly comprehensive model of language-related behaviors would also need additional levels of linguistic (e.g syntactic, phonological) and higher-level cognitive processing. Nevertheless, this work points to a viable framework to aid the development of mutually constraining approaches to robotic communication and to modeling human spatial language processing.

Acknowledgment

The authors acknowledge support from the German Federal Ministry of Education and Research within the National Network Computational Neuroscience - Bernstein Focus: "Learning behavioral models: From human experiment to technical assistance", grant FKZ 01GQ0951.

References

[1] T. Fong, C. Thorpe, and C. Baur, "Collaboration, dialogue, and human-robot interaction," 2001.

[2] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, L. Fadiga, *et al.*, "Integration of Action and Language Knowledge: A Roadmap for Developmental Robotics," *IEEE Transactions on Autonomous Mental Development*, in press.

[3] E. Bicho, L. Louro, and W. Erlhagen, "Integrating verbal and nonverbal communication in a dynamic neural field architecture for human–robot interaction," *Front. Neurorobot*, vol. 4, no. 5, 2010.

[4] N. Mavridis and D. Roy, "Grounded situation models for robots: Where words and percepts meet," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 4690–4697.

[5] V. Krunic, G. Salvi, A. Bernardino, L. Montesano, and J. Santos-Victor, "Affordance based word-to-meaning association," in *ICRA'09: Proceedings of the 2009 IEEE international conference on Robotics and Automation*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 806–811.

[6] D. Skocaj, G. Berginc, B. Ridge, A. Štimec, M. Jogan, O. Vanek, A. Leonardis, M. Hutter, and N. Hawes, "A system for continuous learning of visual concepts," in *International Conference on Computer Vision Systems ICVS*. Citeseer, 2007.

[7] T. Regier and L. Carlson, "Grounding spatial language in perception: An empirical and computational investigation," *Journal of Experimental Psychology: General*, vol. 130, no. 2, pp. 273–298, 2001.

[8] G. Logan, "Linguistic and conceptial control of visual spatial attention," *Cognitive Psychology*, vol. 28, pp. 103–174, 1995.

[9] G. Schöner, "Dynamical systems approaches to cognition," in *Cambridge Handbook of Computational Cognitive Modeling*, R. Sun, Ed. Cambridge, UK: Cambridge University Press, 2008, pp. 101–126.

[10] J. Lipinski, J. P. Spencer, and L. Samuelson, "Towards the integration of linguistic and non-linguistic spatial cognition: A dynamic field theory approach," in *Progress in Neural Processing 18: Proceedings of the Eleventh Neural Computation and Psychology Workshop*, J. Mayor, N. Ruh, and K. Plunkett, Eds. Singapore: World Scientific, 2009.

[11] J. Lipinski, Y. Sandamirskaya, and G. Schöner, "Swing it to the left, swing it to the right: Enacting flexible spatial language using a neurodynamic framework," *Cognitive Neurodynamics, special issue on "Language Dynamics"*, vol. 3, no. 4, 2009.

[12] A. Haasch, S. Hohenner, S. H
"uwel, M. Kleinehagenbrock, S. Lang, I. Toptsis, G. Fink, J. Fritsch, B. Wrede, and G. Sagerer, "Biron–the bielefeld robot companion," in *Proc. Int. Workshop on Advances in Service Robotics*. Citeseer, 2004, pp. 27–32.

[13] D. Roy, K. Hsiao, and N. Mavridis, "Mental imagery for a conversational robot," *IEEE Transactions on Systems, Man, and Cybernetics– Part B: Cybernetics*, vol. 34, no. 3, pp. 1374–1383, 2004.

[14] D. Roy, "Semiotic schemas: A framework for grounding language in the action and perception." *Artificial Intelligence*, vol. 167170-205, pp. 170–205, 2005.

[15] L. Steels and M. Loetzsch, "Perspective alignment in spatial language," in *Spatial language and dialogue*, K. Coventry, T. Tenbrink, and J. Bateman, Eds. Oxford: Oxford University Press, 2008.

[16] L. Steels, "Evolving grounded communication for robots," *TRENDS in Cognitive Sciences*, vol. 7, pp. 308–312, 2003.

[17] W. Erlhagen and G. Schöner, "Dynamic field theory of movement preparation," *Psychological Review*, vol. 109, pp. 545–572, 2002.

[18] S. Amari, "Dynamics of pattern formation in lateral-inhibition type neural fields," *Biological Cybernetics*, vol. 27, pp. 77–87, 1977.

[19] J. Johnson, J. P. Spencer, S. J. Luck, and G. Schöner, "A dynamic neural field model of visual working memory and change detection," *Psychological Science*, in press.

[20] A. R. Schutte, J. P. Spencer, and G. Schöner, "Testing the dynamic field theory: Working memory for locations becomes more spatially precise over development," *Child Development*, vol. 74, no. 5, pp. 1393–1417, 2003.

[21] C. Faubel and G. Schöner, "Learning to recognize objects on the fly: A neurally based dynamic field approach," *Neural Networks*, vol. 21, pp. 562–576, 2008.

[22] W. Erlhagen and E. Bicho, "The dynamic neural field approach to cognitive robotics," *Journal of Neural Engineering*, vol. 3, no. 3, pp. R36–R54, 2006.

[23] Y. Sandamirskaya and G. Schöner, "Dynamic field theory of sequential action: A model and its implementation on an embodied agent," in *7th International Conference on Development and Learning*, 2008.

[24] J. O'Keefe, *Vector grammar, places, and the functional role of the spatial prepositions in English*, E. van der Zee and J. Slack, Eds. Oxford: Oxford University Press., 2003.

[25] I. Iossifidis, C. Bruckhoff, C. Theis, C. Grote, C. Faubel, and G. Schoner, "Cora: An anthropomorphic robot assistant for human environment," in *Proc. 11th IEEE International Workshop on Robot and Human Interactive Communication*, 2002, pp. 392–398.

[26] K. Kreutz-Delgado, Long M, and H. Seraji, "Kinematic analysis of 7-DOF manipulators," *IJRR*, vol. 11, pp. 469–481, 1992.