our future through science

# MODELLING PRIMATE CONTROL OF GRASPING FOR ROBOTICS APPLICATIONS

Ashley Kleinhans[1], Serge Thill[2], Benjamin Rosman[1], Renaud Detry[3] & Bryan Tripp[4]

1.CSIR, South Africa 2.University of Skövde, Sweden

3.University of Liege, Belgium 4.University of Waterloo, Canada

Abstract. The neural circuits that control grasping and perform related visual processing have been studied extensively in Macaque monkeys.We are developing a computational model of this system, in order to better understand its function, and to explore applications to robotics. We recently modelled the neural representation of three-dimensional object shapes, and are currently extending the model to produce hand postures so that it can be tested on a robot. To train the extended model, we are developing a large database of object shapes and corresponding feasible grasps. Finally, further extensions are needed to account for the inuence of higher-level goals on hand posture. This is essential because often the same object must be grasped in dierent ways for dierent purposes. The present paper focuses on a method of incorporating such higher-level goals. A proof-of-concept exhibits several important behaviours, such as choosing from multiple approaches to the same goal.

Keywords: grasping, aordances, macaque, robotics, AIP, F5

## 1 Introduction

The neurophysiology that underlies primate grasping has been studied most extensively in Macaque monkeys. In Macaques, grasping is controlled by an extensive brain network that includes many parts of the visual, parietal, and frontal cortices. A network of dorsal visual and parietal areas detects aordances and may partially parameterize multiple potential movements [1]. Ventral visual and prefrontal areas help to select movements that are consistent with object identities and goals [2]. Our general aim is to translate this rich neurophysiological knowledge into a bio-plausible robotic grasp controller. Specfically, we want to develop a system that uses a robotic hand to grasp a wide range of objects, while reproducing many features of grasp-related neural activity recorded from monkeys.

In pursuit of our goal, we recently developed a neural model [3] that reproduced a variety of electrophysiology data from the caudal and anterior intraparietal areas (CIP and AIP, respectively). These areas encode three-dimensional shape features, and are essential for accurate hand shaping. This model reproduced AIP responses from the Macaque literature using a model of CIP activity as input. We parameterized AIP responses using both superquadric parameters and the parameters of an Isomap reduction of the depth map. We found that both the match with AIP data and the performance of the CIP-AIP mapping were better with Isomap parameters. However, it is not yet clear whether such parameters provide a good basis for grasp planning. For example, in contrast to Isomap, superquadrics support a pose-invariant mapping to some gripper parameters.

To address this question, we have recently started to extend the model to frontal area F5 (which encodes hand postures [4]) so that its applicability to robotic grasp control can be tested. We plan to build a database of grasp examples in order to train and test this extended model. The models trained using such a database will be tested with a real-world robot platform and real objects. We will compare the performance of the neural model to a conventional kernel regression machine, and to state-of-the-art robotics heuristics for grasp planning. We hope to show that a neural model trained on large numbers of examples can provide a practical grasp controller, and that its internal signals are consistent with the literature on neural activity in monkey AIP and F5.

Finally, the main focus of the present paper is on how to further extend the above models to account for how higher-level goals and intentions from prefrontal areas can inuence the decision of which aordances to attend to (and therefore which hand-shape to select). The following sections briey present our approach and a proof-of-concept model. A notable feature of this proof-of-concept is that is expressed entirely in vector operations.

## 2 Methods

Often, different grips are appropriate for manipulating an object for different purposes. For example, if one's goal is to put a hammer in a toolbox, there are many ways in which the hammer can be grasped. However, if the hammer is to be used to hit nails there is essentially one way. To model such influences we are forced to consider a much larger network that includes the prefrontal cortex. The prefrontal cortex is less well understood than the visual cortex, so for these areas the data-driven approach that we previously adopted to model CIP, AIP, and F5 may be less practical. We are instead pursuing a top-down approach based on two key methods. The first is the Neural Engineering Framework [5], which provides a way to map systematically between high-level function and neural activity. The second is Holographic Reduced Representations [6], which are used in cognitive modeling. Recently, these two methods were used together to develop a spiking neural model of the brain with complex cognitive abilities [7]. The methods are described briefly below. For robotics applications, there are various ways to run large models of this type in real time, e.g. surrogate population models on FPGAs [8].

Neural Engineering Framework An NEF model is specied in terms of vector variables that are taken to be encoded by the activity of neuron populations, maps between these vectors, and physiological neuron properties (e.g. time constants). The encoding of a vector by a neural activity is typically modelled as

$$r_i = G\left[\mathbf{e}_i^T \mathbf{x} + b_i\right], \tag{1}$$

where $r_i$ is the spike rate of the $i^{th}$ neuron, $\mathbf{x}$ is the encoded vector, $\mathbf{e}$ is the direction in the encoded space in which the neuron spikes fastest (the "preferred direction"), $b_i$ is a static bias, and $G$ is a physiological nonlinearity. The encoded vector x can be approximately recovered, or "decoded" from the spike rates as

$$\hat{\mathbf{x}} = \sum_i \mathbf{d}_i r_i, \tag{2}$$

where $\mathbf{d_i}$ s called the neuron's "decoding vector", and is chosen to minimize $\mathbf{x} - \hat{\mathbf{x}}$. Furthermore, functions $\hat{\mathbf{f}}(\mathbf{x})$ of the vector can also be decoded by choosing different decoding weights that minimize $\mathbf{f}(\mathbf{x}) - \hat{\mathbf{f}}(\mathbf{x})$. This is the basis of NEF models of neural-network computation. Specifically, if one population encodes $\mathbf{x}$ and a second population encodes $\mathbf{y} = \hat{\mathbf{f}}(\mathbf{x})$, the synaptic weights that produce this mapping can be determined by substituting $\hat{\mathbf{f}}(\mathbf{x})$ ( into (1). The result is that the synaptic weight between the $i^{th}$ presynaptic and $j^{th}$ postsyaptic neuron $w_{ij} = \mathbf{e}_j^T \mathbf{d}_i$. di. Thus, a model can be developed systematically, beginning with a high-level description of encoded variables and how they are transformed.

Holographic Reduced Representations HRRs represent concepts as vectors. They support operations that are useful for cognitive models including binding (associating concepts, e.g. associating "dog" with the role of "actor" in the sentence "dog bites man"); unbinding (e.g. extracting the fact that the "actor" is "dog"), and bundling (combining multiple bound and/or unbound concepts into a package). HRRs use circular convolution for binding and unbinding, and vector addition for bundling. HRR operations are lossy, e.g. "actor" bound to "dog" has the same vector dimension as "actor" or "dog". Eliasmith [9] showed that HRRs can be encoded and manipulated using NEF neural models, and that HRRs of a few hundred dimensions can store tens of thousands of concepts.

### 2.1 Proof-of-Concept Cognitive Model

As a first step in exploring the application of the NEF and HRRs to grasping, we developed a simplified model that uses basic drives and knowledge of the environment to choose a goal, and to influence hand posture in a manner consistent with that goal. To simplify the prototype we used abstract HRR vectors and sigmoidal units, given that the NEF provides a systematic method to develop a spiking neural model from a vector model (this does not work with all vector models, but experience with the NEF suggests that the present model is a good candidate).

Grasping decisions were modelled in the space of the first two principal components of gripper parameters. A grid of sigmoidal units corresponded to different postures in this space. Decisions were made using a diffusion-to-bound mechanism [10], wherein each unit integrates its inputs until one unit's activity crosses a threshold, at which point the winning unit (corresponding to a single posture) inhibits all others. Each input to this network corresponded to the influence of a different brain area on the posture decision, and consisted of a drive pattern across the posture grid. Input from a ten-dimensional object-shape representation was modelled as decoded functions $[f_{ij}(\mathbf{s})]$, where s is the shape parameters and i and j are grid indices. Desired actions were represented in a 200-dimensional HRR. Different actions were nearly orthogonal in this space, so we used a simple linear map, $[\sum_k \alpha_{ijk} \mathbf{a_k}^T \mathbf{a}]$, a, where a are action vectors and $k$ is an index over possible actions.

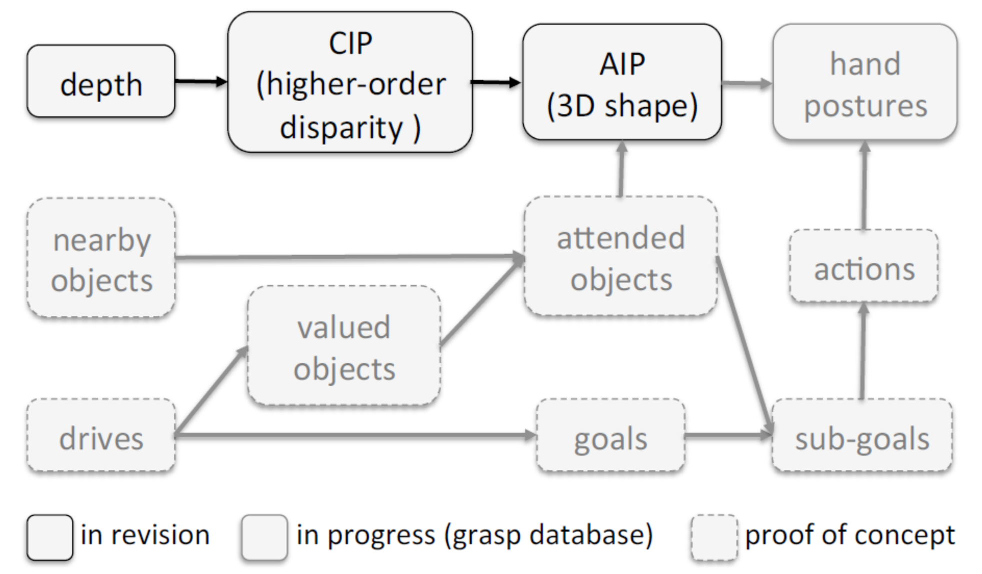


**Fig. 1.** Proof-of-concept model and its relationship to our other work. Dashed boxes indicate HRR populations and a winner take-all "actions" population. Also shown are past work (black boxes) and other current work (solid gray box; see Introduction).

We modelled a scenario in which an agent wants a drink of water given two potential sources: a bottle and a faucet. The agent must decide which source to use and the appropriate hand posture for grasping it. While the scope of this example is somewhat broader than grasp control, we wanted to verify that the basic approach was suitable for such examples. The input to the model included a basic \thirst" drive and a list of the objects in the environment (in a more complete system we take it that these would be detected visually and stored in working memory). We used HRR binding to associate water with both the bottle and the faucet. Furthermore, we used several similar vectors to represent different kinds of water, including cold spring water, warm spring water, and cold tap water. We used linear maps between HRRs to cause a "thirst" concept in the "drives" HRR to probe the "environment" HRR for cold spring water, resulting in selection of the "bottle" concept. Further linear maps between HRRs led to an "action" HRR encoding "grasp" while the "attended object" HRR encoded "bottle". A final linear map from the binding of these two concepts influenced the posture network to choose a posture appropriate for grasping the bottle in order to pour from it.

## 3 Results

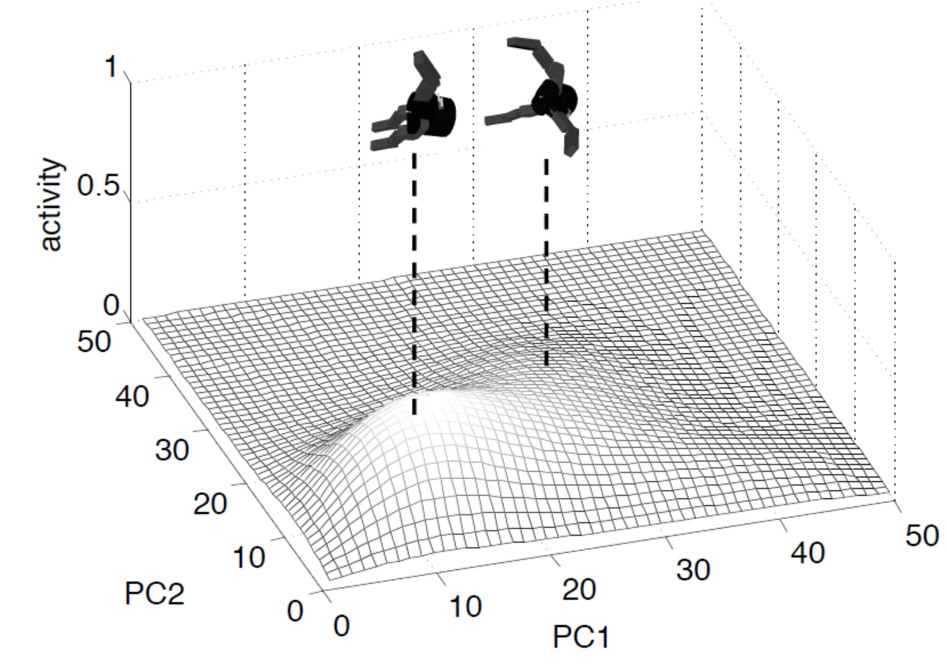


**Fig. 2.** Activation on a grid over the first two principal components of hand posture, during a decision between postures.

Figure 2 shows a snapshot of activity in the hand-posture network, prior to a decision. The insets show two postures of the robot hand that correspond to two potential grips. The one on the left is better suited for lifting the bottle in order to pour from it, and is eventually selected. A different hand posture might be selected if the goal were different (e.g. to put the bottle in a refrigerator) or if the object itself was different.

Simulations of this proof-of-concept model demonstrated promising qualitative properties. First, the model incorporated multiple influences into the selection of a single hand posture. We simulated two specific influences: compatibility with object shape (from AIP); and compatibility with a specified action (from frontal areas). These influences could be arbitrarily broad, narrow, multimodal, etc. Second, the model maps from basic drives to a specific action plan given the objects in the environment. This mapping is oversimplified, but it verifies that such a mapping can be implemented using the NEF and HRRs. Third, the model could choose between multiple routes to the same goal. When we hard-coded the belief that the water bottle was cold, and searched for something similar to cold spring water, attention focused on the bottle. Alternatively, when we hard-coded the belief that the water bottle was warm, attention focused on the faucet instead. We expect that the model could be expanded to include updates based on sensory information.

## 4 Discussion

Two motivations for this research are: curiosity about the primate visuo-motor systems, and practical interest in robot controllers based on the same principles. While similar in spirit to the models studied in robotics [11{17], our work aims to implement affordances, a popular means of formalizing a robotic agent's interaction with the world [18], via a computational model that is compatible with the mechanisms that govern grasping in the primate brain (see [19] for a model with similar goals). In other words, the key novelty is the use of a neurologically plausible model that will nonetheless be implemented on a real robot. Previous robotic implementations tend to at best be cast in connectionist terms inspired by neuroscience (for a discussion, see [18,20]). Models of the relevant brain areas similarly tend to be cast in connectionist terms [19, 21, 22] and analysed for behaviours that resemble that of actual neural circuits. By contrast, the approach discussed in the present paper is closer to biological reality. Although our work is still at an early stage, this gives us hope that we can both achieve more biologically realistic control and contribute to the understanding of biological control mechanisms in a more in-depth manner than connectionist models can.

As an example, let us highlight that we have cast the model first and foremost in terms of a cognitive architecture for which the NEF provides a systematic way of deriving a neural model. As such, this imposes no a priori assumptions on the type and function of neurons in AIP (or F5 for that matter), instead giving us the freedom to investigate the functional contributions of the organisation of these areas [23] directly in terms of a cognitive architecture.

HRRs are a key component of the Spaun model, which can perform a wide variety of sophisticated tasks such as completing patterns from examples. We take the success of this approach in Spaun to suggest that HRRs provide a practical way to integrate a wide range of cognitive inuences (such as verbal instructions) into models of neural visuo-motor systems. Our proof-of-concept model supports this view.

## Acknowledgements

This work was supported by the Swedish Foundation for Strategic Research, the Swedish Research Council, the Belgian National Fund for Scientic Research, a DAAD-NRF Scholarship, and the Natural Sciences and Engineering Research Council of Canada.

## References

1. Fagg, A.H., Arbib, M.A.: Modeling parietalpremotor interactions in primate control of grasping. Neural Networks 11(7-8) (October 1998) 1277{1303
2. Borra, E., Gerbella, M., Rozzi, S., Luppino, G.: Anatomical evidence for the involvement of the macaque ventrolateral prefrontal area 12r in controlling goal-directed actions. 31(34) (2011) 12351{12363
3. Rezai, O., Kleinhans, A., Matallanas, E., Selby, B., Tripp, B.: Hierarchical object representations in the visual cortex and computer vision. Frontiers in Computational Neuroscience (2014) (Accepted). Primate grasping for robotics 7
4. Cerri, G., Shimazu, H., Maier, M.A., Lemon, R.N.: Facilitation from ventral premotor cortex of primary motor cortex outputs to macaque hand muscles. 90(2)(2003) 832{842
5. Eliasmith, C., Anderson, C.: Neural engineering. MIT Press (2003)
6. Plate, T.A.: Holographic Reduced Representation: Distributed representation for cognitive structures. Center for the Study of Language and Inf (2003)
7. Eliasmith, C., Stewart, T.C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., Rasmussen, D.: A large scale model of the functioning brain. Science 338(6111)(November 2012) 1202{1205 PMID: 23197532.
8. Berzish, M., Tripp, B.: A digital hardware design for real-time simulation of large neural-system models in physical settings. CNS (2014)
9. Eliasmith, C.: How to build a brain: A neural architecture for biological cognition. Oxford University Press (2013)
10. Gold, J.I., Shadlen, M.N.: The neural basis of decision making. 30 (2007) 535{574
11. Montesano, L., Lopes, M., Bernardino, A., Santos-Victor, J.: Learning object aordances: From sensory-motor coordination to imitation. IEEE Transactions on Robotics 24(1) (2008) 15{26
12. Stoytchev, A.: Learning the aordances of tools using a behavior-grounded approach. In: Aordance Based Robot Control. Volume 4760 of Lecture Notes in Articial Intelligence (LNAI). Springer (2008) 140{158
13. Sahin, E., Cakmak, M., Dogar, M.R., Ugur, E., Ucoluk, G.: To aord or not to aord: A new formalization of aordances toward aordance-based robot control. Adaptive Behavior (2007)
14. Sun, J., Garibaldi, J.: A novel memetic algorithm for constrained optimization. In: IEEE Congress on Evolutionary Computation. (2010) 1{8
15. Kruger, N., Piater, J., Geib, C., Petrick, R., Steedman, M., Wrgtter, F., Ude, A., Asfour, T., Kraft, D., Omren, D., Agostini, A., Dillmann, R.: Objectaction complexes: Grounded abstractions of sensorymotor processes. Robotics and Autonomous Systems (2011)
16. Detry, R., Baseski, E., Kruger, N., Popovic, M., Touati, Y., Kroemer, O., Peters, J., Piater, J.: Learning object-specic grasp aordance densities. In: IEEE International Conference on Development and Learning. (2009) 1{7
17. Kjellstrom, H., Romero, J., Kragic, D.: Visual object-action recognition: Inferring object aordances from human demonstration. Computer Vision and Image Understanding 115(1) (2011) 81{90
18. Thill, S., Caligiore, D., Borghi, A.M., Ziemke, T., Baldassarre, G.: Theories and computational models of aordance and mirror systems: an integrative review. Neuroscience & BioBehavioral Reviews 37(3) (2013) 491{521
19. Caligiore, D., Borghi, A.M., Parisi, D., Baldassarre, G.: Tropicals: a computational embodied neuroscience model of compatibility eects. Psychological Review 117(4)(2010) 1188
20. Oztop, E., Kawato, M., Arbib, M.A.: Mirror neurons and imitation: A computationally guided review. Neural Networks 19 (2006) 254{271
21. Oztop, E., Imamizu, H., Cheng, G., Kawato, M. Neurocomputing 69(10-12) (June 2006) 1354{1361
22. Thill, S., Svensson, H., Ziemke, T.: Modeling the development of goal-specicity in mirror neurons. Cognitive Computation 3(4) (2011) 525{538
23. Murata, A., Gallese, V., Luppino, G., Kaseda, M., Sakata, H.: Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area aip. 83(5) (2000) 2580{2601 PMID: 10805659.