Project no.:	FP7-ICT-217077
Project full title:	Heterogeneous 3-D Perception across Visual Fragments
Project Acronym:	EYESHOTS
Deliverable no:	D4.2a
Title of the deliverable:	Generating visuo-motor descriptors of reachable objects: Computational approach to the integration problem

Date of Delivery:	08 June 2009
Organization name of lead contractor for this deliverable:	UJI
Author(s):	E. Chinellato, J. B. Grzyb, A. P. del Pobil
Participant(s):	UJI, UG, WWU, UNIBO
Workpackage contributing to the deliverable:	WP4
Nature:	Report
Version:	1.0
Total number of pages:	16
Responsible person:	Angel P. del Pobil
Revised by:	M. Lappe, P. Fattori, F. Hamker, S.P. Sabatini
Start date of project:	1 March 2008 **Duration:** 36 months

Project Co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

This deliverable describes the computational approach to the problem of generating an integrated 3D visuomotor description of the peripersonal space through the interaction with reachable objects by saccades and arm reaching movements toward goal positions, which is the main goal of Task 4.2. We describe the framework of the computational approach we will follow to achieve such goal, an integrated neural representation of the surrounding 3D space built on basis function units. We complete our exposition by presenting part of a data analysis performed on neuroscience studies and describing how the analysis results will be incorporated into the model.

Contents

1 Executive summary

This deliverable describes the research work performed by the Robotic Intelligence Lab of Universitat Jaume I and its partners with regard to the beginning of the second task of Work Package 4 of the EYESHOTS project. The main goal of Task 4.2 is to achieve a visuomotor description of the peripersonal space through active interaction with reachable objects. Such interaction is composed of both saccades and arm movements toward goal positions. The fundamental structure of the description is an integrated neural representation of the surrounding 3D space built on basis function units. This representation is accessed and modified by the possible visuomotor behaviors of the system: saccadic or reaching movements are planned and executed according to this common representation, and the action outcome is used to improve and adjust the representation, and keep the different visuomotor components in register with respect to each other and to the actual layout of the peripersonal space.

After the introduction, this report presents a background section (Section 3) which includes a description of the neuroscience concepts related to visuomotor interactions with the objects in the peripersonal space in human and other primates, and a summary of the computational characteristics of the basis function approach to sensorimotor transformations. In Section 4, the proposed computational approach is described with more details, highlighting its novelties compared to previous models, and the relation with experimental data collected by other EYESHOTS partners on humans and monkeys, which analysis is partially described in the same section.

2 Introduction

Humans and other primates build their perception of the surrounding space mainly through active interaction with nearby objects that allows them to build a representation of the environment useful for further interactions. The main sources of knowledge that are used to build such representation are retinotopic (visual information) and proprioceptive. The dorsal visual stream of the primate brain, and more precisely the posterior parietal cortex (PPC) is very likely to contain the areas in charge of accessing and updating the representation. Areas within the PPC are in charge of performing the reference frames transformations required to map visual information to appropriate oculomotor and limb movements. It is often argued, and increasingly accepted by the neuroscientific community, that such ability is achieved through the use of gain fields and basis function representations, that allow us to simultaneously represent stimuli in various reference frames, rather than on a set of different reference frames that are all maintained in memory and sequentially accessed through a cascade of transformations. Such a basis function representation will be the core of the model that we are presenting. We will build it by relying upon findings from human and monkey studies, and especially from data on human saccade and arm movement provided by the WWU partner and the modulation between visual information, gaze direction and arm reaching movements in monkey posterior parietal area V6A provided by the UNIBO partner.

For what concerns the type of visuomotor knowledge of the environment required to our system, this should finally be able to construct a representation of its peripersonal space in a dynamical way, through the practical interaction with the environment features, using both visual input and proprioceptive data concerning eye and arm movements. Following this approach, the robotic arm should naturally achieve very good open-loop reaching and saccade capabilities towards known targets. This goal is represented in Figure 1, which depicts a simple conceptual schema of how the space representation is generated and updated. Peripersonal space is represented by a plastic map, that constitutes both a knowledge of the environment and a sensorimotor code for performing movements and evaluate their outcome. The map is accessed and modified by two types of information: retinotopic (visual) and proprioceptive (eye and arm movements). Contextually, eye and arm motor plans are devised in accordance to the map itself.

The exploration of the environment through saccades and reaching movements constitutes the basic behavior that is employed to build the visuomotor representation of the peripersonal space. As a first step in the implementation and the experimental development, the perception of the space and the related sensorimotor map will be accessed and updated by only one visuomotor interaction, e.g. move the gaze and the arm toward a goal position. As a long-term objective, the global, integrated representation of the 3D space will be used to process sequences of events and there will be the need to keep track of previously executed actions. If the sequence of actions involves some topological representation of the peripersonal space, and memorized objects and locations, this would involve aspects related to the perception-oriented ventral stream, which is in charge of coding objects in the visual space relatively to each other, and to keep track of "historic" sequences of events. Thus, as a final goal the agent should be able to purposefully build such visuomotor map of the environment in 3D from scratch, simultaneously learning to look at and reach towards different visual targets, and achieving binding capabilities through active exploration.
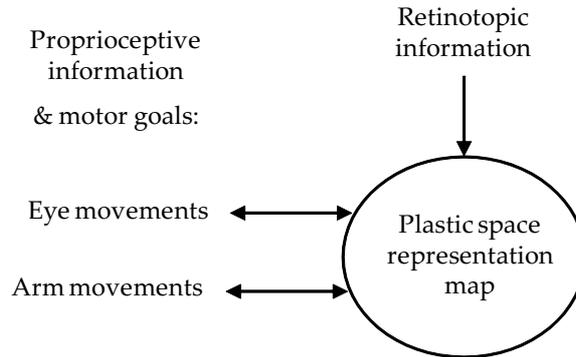
Figure 1: Conceptual schema of the final space representation map

3 Background

The visual cortex of the primate brain is organized in two parallel channels, called "dorsal" and "ventral" streams. The former elaborates visual data with the main purpose of endowing the subject with the ability of interacting with his/her environment, and its tasks are often synthesized as "vision for action". The latter is dedicated to object recognition and conceptual processing, and thus performs "vision for perception". Although a tight interaction between the two streams is necessary for most everyday tasks, dorsal stream areas are more strictly related to the planning and monitoring of reaching and grasping actions [15]. In fact, dorsal visual analysis is driven by the absolute dimension and location of target objects, requiring continuous transformations from retinal data to an effector-based frame of reference. Such frame of reference is body-centered for arm movements (or body-centered, since we assume the head to be fixed) and retinocentric for gaze movements. The ventral stream maintains instead a contextual coding of objects in the environment, based on their identity and meaning. Spatially, such coding can be defined as object-centered, as it is mainly concerned with the relative location of objects with respect to each other.

The correct coupling between the reaching and grasping movements, often neglected in robotic applications, is instead a fundamental and largely studied aspect in human grasping, and various plausible models on the relation between reaching and preshaping movements have been developed [25]. The hypothesis of parallel visuomotor channels for the transport and the preshaping components of the reach-to-grasp action is well recognized [17]. Anatomically, these two channels fall both inside the dorsal stream, and are sometimes named dorso-medial and dorso-lateral visuomotor channels [14]. Cortical areas nomenclature is still controversial, and the correspondence between human and macaque studies not completely solved, but new studies confirm the duality of the reaching-grasping process [9]. For what concerns proximal joint movements, focus of interest of this research, and according to a well established nomenclature, the most important reach-related cortical areas are V6A and MIP, both receiving their main input from V6 and projecting to the dorsal premotor cortex [14, 12, 10].

3.1 Sensorimotor transformations in the posterior parietal cortex

In order to elaborate a proper action on an external target, the dorsal stream requires two main inputs, the object shape and pose and its location with respect to the eyes and to the hand. These inputs are obtained by integrating retinal information regarding the object with

4

proprioceptive data referred to eyes, head and hand. All this information is managed contextually by the dorsal stream, through its two parallel sub-streams, dorso-medial and dorso-lateral. Area V6A seems to represent a fundamental relay station in this complex network. The assumption is that information regarding eye position and gaze direction is employed by V6A in order to estimate the position of surrounding objects and guide reaching movements toward them. Two types of neurons have been found in V6A that allow us to sustain this hypothesis [13]. The receptive fields of neurons of the first type are organized in retinotopic coordinates, but they can encode spatial locations thanks to gaze modulation. The receptive fields of the second type of neurons are organized according to the real, absolute distribution of the subject peripersonal space. In addition, V6A contains neurons that arguably represent the target of reaching retinocentrically, and others that use a spatial representation [19]. This strongly suggests a critical role of V6A in the gradual transformation from a retinotopic to an effector-centered frame of reference. Moreover, some V6A neurons appear to be directly involved in the execution of reaching movements [14], indicating that this area is in charge (probably together with MIP) of performing the visuomotor transformations required for the purposive control of proximal arm joints, integrating visual, somatosensory and somatomotor signals in order to reach a given target in the 3D space.

In our framework, eyes and arms are treated as separate effectors that receive motor control via different specific representations which, however, combine to form a unique shared representation of visuo-motor awareness. The cortical representations that control eye movements, such as area LIP in the parietal cortex and area FEF in the frontal cortex, use a retino-centric reference frame. However, in the basis function framework, explicit encoding of targets in retino-centric coordinates might be enhanced via gain fields to hold in parallel an implicit encoding in other reference frames. Such gain fields are found in retino-centric organized eye movement areas LIP [1, 3] and FEF [5].

The proposed combination of eye and arm movement control into a common representation of spatial awareness relies on the fact that eye and arm movements usually go together, i.e., we fixate an object before, or while, we reach towards it. Such combinations of looking and reaching for the same target can be used to establish a common representation by observing the errors of the gaze and of the reach movement and using learning to bring both errors to zero and thus both representations in register. Experimental support for both the plasticity of the spatial awareness and the conjunction between gaze and arm movements in this plasticity comes from studies of saccadic adaptation and its transfer to arm movements. In saccadic adaptation, an artificial error of saccade targeting is introduced experimentally by displacing the target of a saccade during the saccade execution [20]. This error gives rise to an adaptation procedure that adjusts the parameters of the saccade to reach the new, displaced target position. Part of this adjustment results from an adaptation of the spatial representation of the target, i.e. involves a modification of the sensorimotor map [2, 7, 11]. Parter WWU is actively investigating the visual and motor properties of such adaptation in the framework of Task 5.3 [18].

This modification transfers to reaching and pointing movements [8, 4, 16], clearly showing that the adaptation takes places in a joint representation. This mechanism allows us to keep both eye and arm targeting in register and to establish a common spatial awareness of the target's location. Our framework is therefore based on a common code for spatial awareness obtained by a learning procedure based on target errors of eye or arm movements to visual targets, as described in Section 4.

5

3.2 The basis function approach to sensorimotor transformations

Basis functions are building blocks that, when combined linearly, can approximate any non-linear function, such as those required to map between different neural representations of the peripersonal space (retinotopic, head-centered, arm-centered). Basis function networks have been proposed as a computational solution especially suitable to model the kind of sensorimotor transformations performed by the posterior parietal cortex [21, 22, 23]. Networks of suitable basis functions are in fact able to naturally reproduce the gain-field effects often observed in parietal neurons [24]. Pouget and Sejnowski [21] suggested that positions of object in the peripersonal space are coded through the activity of parietal neurons that act as basis functions, and any coordinate frame can be read out from such population coding according to the task requirements.

Several different transfer functions can be used as basis functions, the only requirements are that they are non-linear, that their interaction is also non-linear (e.g. product vs. sum), and that they cover all the possible input range. The most used functions, for their convenience and biological plausibility, are Gaussian and sigmoid functions. For example, retinotopic maps are often modeled by Gaussian basis functions, and eye position by sigmoid, or logistic, functions [21]. Learning in basis function networks is composed of two stages, the first for choosing the shape and location of the basis functions and the second to map them to the output representation. The first step is unsupervised, the second depend on errors observed during the sensorimotor interaction with the world. Short-term or working memory of up to a few seconds can be introduced in basis function networks through lateral connectivity [22].

4 Outline of the computational approach

As mentioned in Section 2 and considering the theoretical description of Section 3, the main source of inspiration for our model is the basis function approach [21, 24]. The more relevant components of the model are reproduced as a block diagram in Fig. 2 [23]. Several modifications and extensions are nevertheless necessary in order to adapt such approach to our current requirements.

The building blocks of our framework are shown in Fig. 3. First of all, the quality and the nature of the sensory stimuli to be introduced is different. The process of 3D localization requires the integration of information coming from various sources and different modalities. The modeling of such process can be done at different levels of detail and considering alternative data sources and formats. In our case, we include visual information and data on eye position and arm position. Moreover, as a fundamental novelty and improvement, binocular information is introduced in the model instead of the usual monocular one. Several possible alternatives for representing this information can be employed. One natural solution is to include both left and right retinotopic maps, and also left and right eye gaze directions. Nevertheless, considering that we are modeling extrastriate and associative visual areas, it is plausible to assume that stereoptic information is represented, on the proprioceptive part, by version and vergence angles (in the schema, they would substitute the two eye positions). On the visual part, an alternative to the two retinotopic maps is the composition of a cyclopean image representation with a disparity map (under the assumption that the correspondence problem is already solved). On our view, the same transformation that is used to recode possible targets in a body-centered reference frame can be used to extract a 3D cyclopean map from binocular data. At the other hand of the transformation scheme (it is not an

output, since basis function representations do not impose a preferential processing direction, but allow instead for bidirectional transformations), we find the body-centered map of the peripersonal space (which is also head-centered in our case). This scheme allows us to transform ocular movements and stereoptic visual information to a body-centered reference frame but also, when needed, elicit the eye movements that are necessary to foveate on a given visual target. The map is thus accessed and updated upon requirements, as described in the next subsection.
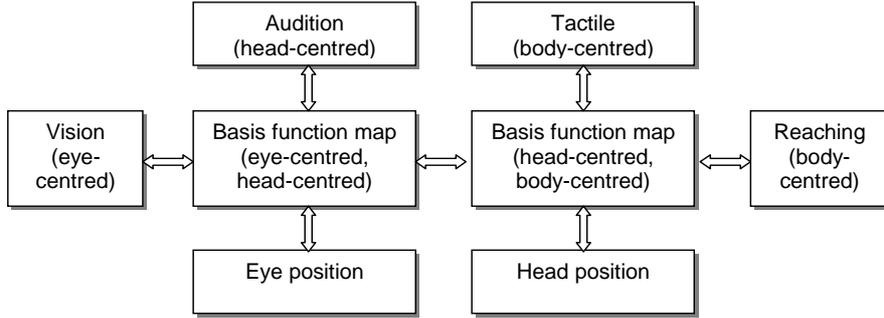
Figure 2: Framework proposed by *Pouget et al.* [23]
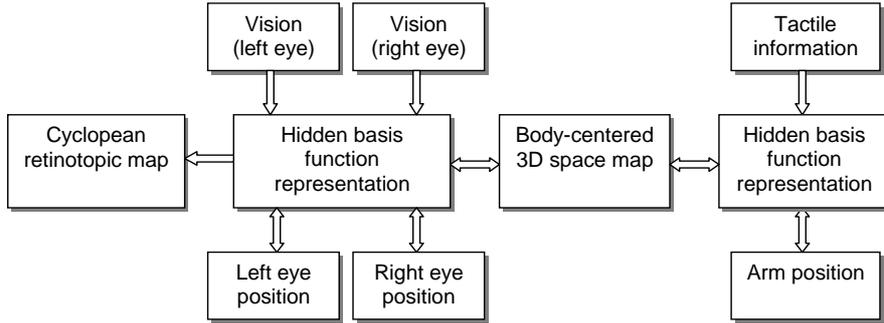
Figure 3: Building blocks of the global 3D space representation. The body-centered 3D map constitutes the integrated visuomotor representation of the peripersonal space.

The basis function approach has the attractive feature that both head-centric representations for arm movements and retino-centric representations for gaze movements are encoded concurrently in the same neural map. As a second extension, the same map will thus be used also for coding the position of objects in the peripersonal space considered as potential targets for reaching movements. In this case, proprioceptive information regarding arm position is included in the computation, and the vectors corresponding to reaching movements can be extracted similarly to what is done for ocular movements. Again, upon tactile feedback confirming that the target object has been reached, the actual reaching direction can be compared with the expected one and the 3D visuomotor representation accordingly updated if necessary.

The third fundamental improvement we will introduce is the consolidation of the sensorimotor map required for executing saccades or reaching movements toward remembered targets. This capability can be achieved only employing some memory-based mechanisms. It has been mentioned above that short-term, working memory can be achieved through the

insertion of lateral connection between the basis function units. More elaborate coding of visuomotor events would nevertheless require other mechanisms, such as saliency maps to highlight "interesting" positions in the peripersonal space. In any case, any change in the appearance of the visually reachable space should elicit a double movement, saccade and reach, toward the focus of attention constituted by the "mutating" position.

4.1 Getting to know the environment

It is worth reminding that the proposed computational scheme is designed to be actually applied to a real robotic setup. The final goal is to provide the robot with advanced skills in its interaction with the environment, namely in the purposeful exploration of the peripersonal space and the contextual coding and control of eye and arm movements. The use of a robot hardware constitutes a possible complication in the realization of the model, and some issues that would easily be solved in simulated environments have to be dealt with more accurately considering the real world implementation. On the other hand, the implementation on an actual sensorimotor setup is a potential source of additional insights hardly achievable with simulated data. Moreover, to the best of our knowledge, this is the first research that applies the basis function approach to robotic oculomotor and reaching control.

Although in principle one representation should be enough for all the required transformations, the number of neurons required to contextually code for n different signals is given by the size of the signals to the power of n. It is easy to see that a representation maintaining both eye visual and proprioceptive signals, and arm joint information would be computationally unfeasible, even for the brain itself. A more logical structure is one in which a central body-centered representation is accessed and updated both by limb sensorimotor signals on the one hand and visual and oculomotor signals on the other hand. Indeed, this seems to be how the problem is solved within the brain, in which different areas or populations of neurons in the same areas are dedicated to different transformations. Most importantly, this approach is consistent with the findings related to area V6A, which contains neurons that code for the retinocentric position of targets, others that code for their limb-related position, and even others that seem to maintain both codings and thus perform the transformation.

The global structure of the model will thus be modular, at least separated in the visual and oculomotor component and the arm sensorimotor component (left and right of Fig. 3). This has also the advantage of allowing us to check for the achievement of the different capabilities separately, and simplify the first part of the learning process. In fact, the first learning stage consists in establishing the structure of the hidden layers composed by non-linear basis functions, that can be shaped as proposed in previous models, product of Gaussians and sigmoids, but also differently if required. In any case, there will always be a common representation that is accessed and modified by a conjunction of gaze and reach movements to the same target. The second learning stage builds on this principle. In this stage, the actual map of the peripersonal space is learnt through active exploration, and the weights of the basis functions that associate the input patterns to the internal representation are dynamically learnt and updated whenever needed. This second part of the learning process is indeed part of the normal behavior of the agent, and constitutes the most fundamental component of its basic capability of interacting with the world and contextually update its representation of the world itself. Such goal can be achieved by adjusting the weights of the basis functions according to the errors observed in the oculomotor and reaching movements as guided by the same basis functions. The concurrence of eye and arm movements should be sufficient

to provide the appropriate error signals even when actual distances between target and final position of the action are not explicitly provided. We could call this approach as a "self-supervised learning" framework, in which the different modalities supervise each other, and eye and arm movements both improve, and obtain together a precise visuomotor representation of the surrounding space. The presence of a tactile response as feedback for the correct execution of a reaching movement is critical to provide a reliable "master" signal to ensure the accuracy of the global representation.

The inclusion of memory mechanisms in the framework, already mentioned above, is an issue that will need to be solved during the implementation of the model. For what concerns saccade to recently observed targets, this should be made possible by the use of lateral connections. Regarding more complex topological representations of the peripersonal space, we are working on a simple object identification framework that models a dorsal contribution to the job of the ventral stream. Such approach is based on the employment of grasping-oriented object analysis – namely, the SOS/AOS coding of neurons in the posterior intraparietal sulcus CIP – for recognition purposes in the lateral occipital complex LOC [6].

4.2   The different aspects of neural response during reaching

The last aspect that is fundamental for our modeling purposes is the use of insights and functional indications coming from the data analysis we are performing in collaboration with the WWU and UNIBO partners, regarding humans saccadic movements and the transformations and the contextual encoding of features in the peripersonal space performed by area V6A. The use of quantitative experimental data regarding neural activation in the model is a distinguishing feature that endows our approach with a special experimental stance, which is missing even in well recognized models. The fundamental results of our analysis of the UNIBO partner data and the way we will use them in the model are described in the following.

In previous works, single-cell experiments performed on macaque monkeys were described and analyzed [12, 13, 19]. Here, we aim at shedding further light on the sort of transformations performed by V6A neurons and on the coding representations they use for this purpose. The analysis approach we employ has the final goal of achieving a computational description of V6A neurons to be used within the above described computational approach.

Experiment Description

The experiments analyzed here were collected at the Università di Bologna on two trained macaque monkeys. They were approved by the Bioethical Committee of the University and carried out in accordance with Italian national laws and European Directives on care and use of laboratory animals. Data were collected while the monkeys were performing two possible reaching tasks to given targets while gazing at a certain position (the fixation point) illuminated by an LED (Figure 4). In the first task (**Constant reaching**) the target remained always in the same straight-ahead position, whereas the fixation point could be in one out of three different positions, as symbolized in Figure 4(a). In the second task (**Foveal reaching**) the fixation point changed in one out of three positions as in the first task, but arm-reaching movements were always directed towards the fixation point, as depicted in Figure 4(b). For other details regarding experimental procedures see [19].

The data analysis focuses on the average neural firing rate during four time intervals of the action course (epochs). The time epochs taken into account were defined as follows:

(a) Constant reaching protocol

(b) Foveal reaching protocol
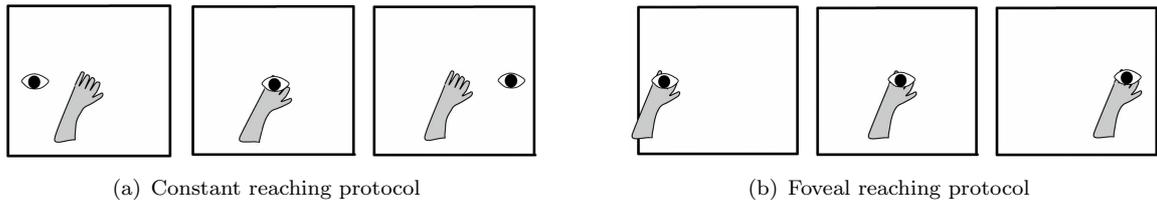
Figure 4: Graphical description of experimental protocols.

FIX: steady fixation of the LED; it starts when the gazing on the fixation point is detected and ends at the onset of the position cue indicating the position to be reached;

DELAY: delay period before the go-signal; starts 300ms after the position cue offset and ends at the go-signal.

MOV: arm reaching movement; starts 200ms before movement onset and lasts until movement end.

HOLD: object holding period; starts at movement end and finishes 200 ms before return movement onset.

Two main types of analysis were performed on the data, one based on the preferred response of neurons, the other on a principal component analysis of their responsiveness.

Preferred Direction and Temporal Properties of the Neural Response

The first step of this analysis was to compute for all neurons a preferred direction index, in the two protocols and for each epoch of interest. This was done by calculating an average of the three possible positions weighted by their firing rates. The responsiveness of each neuron was thus expressed by 8 values: its preferred direction in each of the 4 epochs of interest for both Constant and Foveal reaching protocols.

As a first step, we observed that the responsiveness of the neural population spans the entire working range, and that neurons preferred directions assume an approximately Gaussian distribution symmetrical with respect to the central direction, as it can be observed in the example of Figure 5, referred to the Constant reaching protocol. It remains to be verified how the choice of the target positions affect such distribution, and it cannot be excluded that other neurons would be selective for positions further away from the center. In any case, the fact that all input space is covered by the neural population satisfies one of the basic requirements of basis function structures.

(a) FIX epoch

(b) DELAY epoch

(c) MOV epoch

(d) HOLD epoch
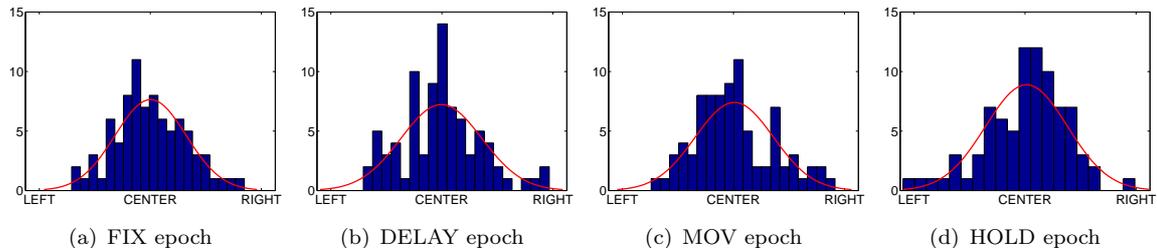
Figure 5: Preferred direction: within epoch distributions.

10

Table 1: Correlation Coefficient between different epochs for constant and foveal reaching

Epochs	Constant reaching	Foveal reaching
FIX/ DELAY	0.15	0.90
FIX/ MOV	0.02	0.76
FIX/ HOLD	0.20	0.61
DELAY/ MOV	0.86	0.83
DELAY/ HOLD	0.56	0.81
HOLD/ MOV	0.54	0.92

More interesting insights can be drawn from a comparative assessment of neurons preferred directions in different conditions and epochs. The results obtained comparing the preferred directions of neurons during the same epochs in the two experimental tasks are depicted in Figure 6 for epochs FIX and MOV. Epoch FIX shows the highest correlation between protocols ($R = 0.69$), indicating a consistent activation across tasks during this epoch (Figure 6(a)). On the other hand, no correlation between tasks was found for the MOV epoch ($R = -0.02$, Figure 6(b)). Correlation coefficients for DELAY and HOLD tasks are $R = 0.36$ and $R = 0.24$, indicating some correlation, but not a strong one. These results suggest that the change in protocol affects principally the motor components of the neural responsiveness, while gaze selectivity (mainly referred to epoch FIX) is largely unaffected by the movement change. The mixed neuronal response elicited by DELAY and HOLD epochs may indicate a dual nature of this processing steps, composed of both visual and motor components. Possible correlations are more apparent if only neurons selective in one or both tasks are considered (see color-coding in Figure 6).
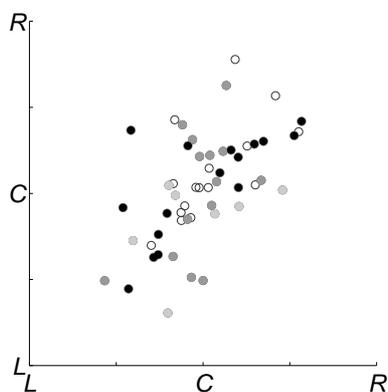
(a) FIX epoch: Constant (x) vs. Foveal (y)   (b) MOV epoch: Constant (x) vs. Foveal (y)
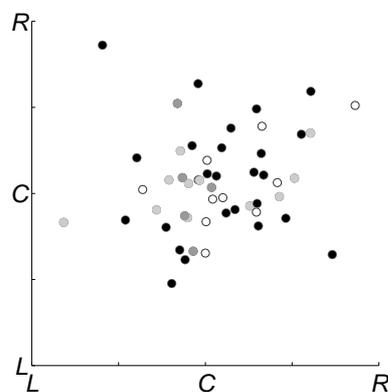
Figure 6: Preferred direction: same epoch, different tasks (L=left; C=center; R=right). Dot color = neuron selectivity: white - not selective; light gray - selective in Constant; dark gray - selective in Foveal; black - selective in Constant and Foveal.

Relevant considerations can be drawn also by the study of how neural responsiveness changes during the action course within the same experimental protocol. This can be done comparing the preferred direction of neurons in the same task but in different epochs. The correlation coefficients we obtained in this analysis are summarized in table 1. The high-

est and most consistent correspondence in preferred directions can be observed between the DELAY and MOV epochs for both Constant and Foveal protocols, suggesting a certain processing uniformity across such epochs. No correlations are found between FIX and other epochs for the Constant protocol, and only mild correlations are observed in DELAY/HOLD and HOLD/MOV. In Foveal reaching the situation is different, as all epochs show some correspondence, and especially the three epochs DELAY-MOV-HOLD are very well correlated, but also FIX is now much more correlated with all other epochs. This could indicate than, when the gaze is directed where the hand is (Foveal reaching) there is a coupling in the discharge in HOLD and the epochs preceding it. Conversely, when the hand is maintained in a location not gazed at (Constant reaching), the cell discharge can be uncorrelated to DELAY and MOV activity probably because the spatial coordinates used in that stage are in a different frame of reference. Looking for a correspond with our schema of Fig. 3, we could hazard that the FIX epoch corresponds mainly with the left part of the framework and the MOV epoch with the right part, while activation might be more distributed for the DELAY and HOLD epochs.

As a general insight, some neurons seem to maintain their responsiveness across epochs and protocols, others completely change their preferred direction. These findings suggests the presence of important temporal issues, and a strong effect of action stage on neural responsiveness. A possible interpretation of this activity pattern is that some neurons sustain their activation, maybe for maintaining their coding of the target position in the head-centered reference frame, whereas others perform transformations between this frame and the retinocentric and arm-centered references. The first type of neurons would be in charge of maintaining the common visuomotor spatial representation in our model, whilst the other neurons would be in charge of accessing and modifying such representation according to different sensorimotor events.

Principal Component Analysis

In order to better understand the sort of representation used by V6A neurons, the next step in our analysis was to perform a principal component analysis of the responsiveness of all neurons and conditions (LEFT, CENTER, RIGHT) of an experimental protocol for each epoch of interest. PCA was thus executed over a 87x3 dataset for each epoch, and in all cases, the first two principal components accounted for nearly or more than 90% of the data variability. Thus, for both Constant and Foveal reaching, two components are almost enough to represent the whole range of the three different experimental conditions. This means that most neurons are "predictable" in their activity pattern, showing reasonably monotonic activation patterns. A normalized representation of the three eigenvectors obtained for each epoch during Constant and Foveal reaching is depicted in Figure 7. The relative weights of the eigenvectors, which exemplify their capacity of representing the whole dataset, and obtained normalizing their eigenvalues, are also provided.

A first interesting aspect that can be noticed is the strict similarity between the principal components of the DELAY and the MOV epochs (Figures 7(b) and 7(c)). Such finding confirms and reinforces the hypothesis that neuronal activation between these two epochs is very consistent. In Constant reaching, a very good correspondence can also be observed between the FIX and HOLD epochs (Figures 7(a) and 7(d)), showing a relation between them that do not correspond to their low correlation. For the Foveal reaching protocol (Figures 7(e-h)), one major change is noticeable with respect to Constant reaching: while the correspondence between DELAY and MOV remains clear, epoch HOLD is now definitely
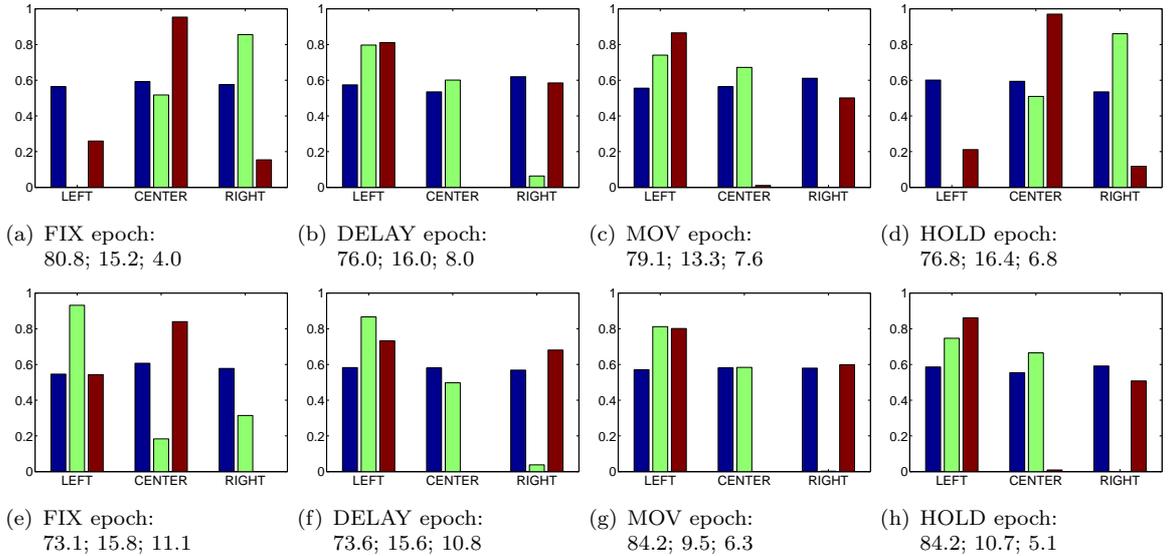
Figure 7: PCA for Constant (above) and Foveal reaching (below). Principal components of each epoch across conditions, with correspondent weights (%).

closer in its principal components to these two epochs than to FIX. It is also interesting to observe how DELAY and MOV principal components remain consistent across protocols. The correspondence between the HOLD and DELAY/MOV epochs in the Foveal task and not in the Constant reaching task could be explained considering that in the first case the attention of the subject is directed toward the same position during DELAY (while planning the movement), MOV (while executing the movement), and HOLD (while waiting for the Return signal). In the second task, instead, this correspondence is present for DELAY and MOV, but not for HOLD. Indeed, in the latter epoch the subject is holding its hand in one position, but its visual attention is directed toward the fixation point.

From a computational point of view, there are two ways to bridge the gap between the model and the quantitative results obtained from the experimental data. The first, common method is to define the structure of the artificial neural population according to the property the entire population is expected to possess. Then, upon training with real experiments, the behavior and coding of artificial neurons have to be compared to that of the real primate neurons.

An alternative approach that we will explore is to employ the principal components obtained in this analysis as a first approximation for modeling the functional behavior of V6A neurons. The population of neurons that constitute the model can be generated taking into account such components considering them as fundamental basis functions. The different relations between tasks and epochs observed in the analysis would bias the composition and behavior of the populations of neurons in the model.

In general, the different properties captured in this analysis can be used to tune the behavior of the neural population with various input sets corresponding to the different experimental conditions. More experiments concerning the execution of similar movements in light and dark conditions are being analyzed and will possibly also be taken into account in the model.

5 Conclusions

A robot emulating the processing of the primate dorsal visual stream should be able to purposefully and consistently interact with its environment building its skills on the integration of different stimuli. Such skills would be based on the building of a plastic representation of its nearby environment, representation that can be exploited for more precise and complex interactions with the environment components. In this deliverable, we described the framework we propose for implementing, first computationally and later on a real robotic setup, the ability of safely and reliably interacting with objects in the peripersonal space through the construction, access and contextual updating of an integrated visuomotor 3D map of the surrounding environment, based on basis function units.

We also described part of our data analysis aimed at a better understanding the role of the dorso-medial visual stream in the planning and execution of reaching actions, which helps in clarifying what sort of computation is performed by dorsal stream neurons, namely those pertaining to area V6A, in order to maintain a perfect coordination between retinal data, gaze direction and arm movements. We explained how we will use such findings to bias the implementation of the proposed model framework, which constitutes the next step of our research within the EYESHOTS project.

Robotic experiments will finally help in further clarifying the mechanisms behind eye-arm coordination and reciprocal guidance and reference frame transformations in primates. A first interesting test is to extend the one-dimensional nature of the experiments presented in this work first to 2D and then, to the full 3D space, adding depth information processing and check how the mutual modulation between retinal data, gaze direction and reaching movements is required to change to adapt to the different cases. This should carry to a better understanding of the transformations performed between retinocentric, effector-based and distance/vergence-based representations in various environments and working conditions. The predictions obtained by the model and the robotic experiments could then be tested through the development of new neuroscience studies.

Bibliography

[1] R. A. Andersen, R. M. Bracewell, S. Barash, J. W. Gnadt, and L. Fogassi. Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7A of macaque. *J. Neurosci.*, 10(4):1176–1196, 1990.

[2] H. Awater, D. Burr, M. Lappe, M. C. Morrone, and M. E. Goldberg. Effect of saccadic adaptation on localization of visual targets. *J. Neurophysiol.*, 93(6):3605–3614, 2005.

[3] F. Bremmer, C. Distler, and K.-P. Hoffmann. Eye position effects in monkey cortex. II: Pursuit and fixation related activity in posterior parietal areas LIP and 7A. *J. Neurophysiol.*, 77:962–977, 1997.

[4] A. Bruno and M. C. Morrone. Influence of saccadic adaptation on spatial localization: Comparison of verbal and pointing reports. *J. Vis.*, 7 (5)(16):1–12, 2007.

[5] C. R. Cassanello, A. T. Nihalani, and V. P. Ferrera. Neuronal responses to moving targets in monkey frontal eye fields. *J. Neurophysiol.*, 100:1544–1556, 2008.

[6] E. Chinellato and A. P. del Pobil. Neural coding in the dorsal visual stream. In *International Conference on the Simulation of Adaptive Behavior*, 2008.

[7] T. Collins, K. Dore-Mazars, and M. Lappe. Motor space structures perceptual space: Evidence from human saccadic adaptation. *Brain Res.*, 1172:32–39, August 2007.

[8] J. Cotti, A. Guillaume, N. Alahyane, D. Pelisson, and J.-L. Vercher. Adaptation of voluntary saccades, but not of reactive saccades, transfers to hand pointing movements. *J. Neurophysiol.*, 98(2):602–612, 2007.

[9] J. C. Culham, J. P. Gallivan, C. Cavina-Pratesi, and D. J. Quinlan. fMRI investigations of reaching and ego space in human superior parieto-occipital cortex. In M. Behrmann & B. MacWhinney R. Klatzky, editor, *Embodiment, Ego-space and Action*, pages 247–274. Lawrence Erlbaum Associates, 2008.

[10] P. Dechent and J. Frahm. Characterization of the human visual V6 complex by functional magnetic resonance imaging. *Eur J Neurosci*, 17(10):2201–2211, 2003.

[11] V. Ethier, D. S. Zee, and R. Shadmehr. Changes in control of saccades during gain adaptation. *J. Neurosci.*, 28(51):13929–13937, Dec 2008.

[12] P. Fattori, M. Gamberini, D. F. Kutz, and C. Galletti. 'Arm-reaching' neurons in the parietal area V6A of the macaque monkey. *Eur J Neurosci*, 13(12):2309–2313, Jun 2001.

[13] P. Fattori, D. F. Kutz, R. Breveglieri, N. Marzocchi, and C. Galletti. Spatial tuning of reaching activity in the medial parieto-occipital cortex (area v6a) of macaque monkey. *Eur J Neurosci*, 22(4):956–972, Aug 2005.

[14] C. Galletti, D. F. Kutz, M. Gamberini, R. Breveglieri, and P. Fattori. Role of the medial parieto-occipital cortex in the control of reaching and grasping movements. *Experimental Brain Research*, 153(2):158–170, November 2003.

[15] M. A. Goodale and A. D. Milner. *Sight Unseen.* Oxford University Press, 2004.

[16] T. D. Hernandez, C. A. Levitan, M. S. Banks, and C. M. Schor. How does saccade adaptation affect visual perception? *J. Vis.*, 8 (8)(3):1–16, 2008.

[17] M. Jeannerod. Visuomotor channels: Their integration in goal-directed prehension. *Human Movement Science*, 18(2):201–218, June 1999.

[18] M. Lappe. What is adapted in saccadic adaptation? *J. Physiol.*, 587(Pt 1):5, 2009.

[19] N. Marzocchi, R. Breveglieri, C. Galletti, and P. Fattori. Reaching activity in parietal area V6A of macaque: eye influence on arm activity or retinocentric coding of reaching movements? *Eur J Neurosci*, 27(3):775–789, Feb 2008.

[20] S. McLaughlin. Parametric adjustment in saccadic eye movements. *Percept. Psychophys.*, 2:359–362, 1967.

[21] A. Pouget and T. J. Sejnowski. A new view of hemineglect based on the response properties of parietal neurones. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352(1360):1449–1459, October 1997.

[22] A. Pouget and L. H. Snyder. Computational approaches to sensorimotor transformations. *Nature Neuroscience*, 3 Suppl:1192–1198, November 2000.

[23] Alexandre Pouget, Sophie Deneve, and Jean-Ren Duhamel. A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, 3(9):741–747, September 2002.

[24] E. Salinas and P. Thier. Gain modulation: a major computational principle of the central nervous system. *Neuron*, 27(1):15–21, July 2000.

[25] R. Shadmehr and S. P. Wise. *The computational neurobiology of reaching and pointing: A foundation for motor learning.* MIT Press, 2005.