PROJECT PERIODIC REPORT

Grant Agreement number: 217077

Project acronym: EYESHOTS

Project title: Heterogeneous 3-D Perception across Visual Fragments

Funding Scheme: Collaborative project

Date of latest version of Annex I against which the assessment will be made: 05/10/2007

Periodic report: 1st ☐ 2nd ☒ 3rd ☐

Period covered: from 01/03/2009 to 28/02/2010

Name, title and organisation of the scientific representative of the project's coordinator:

Silvio P. Sabatini, Dr.

Department of Biophysical and Electronic Engineering - DIBE

University of Genoa - UG

Tel: +39-010-353-2092 (2289)

E-mail: silvio.sabatini@unige.it

www.eyeshots.it

TABLE OF CONTENTS

Declaration by the scientific representative of the project coordinator

I, as scientific representative of the coordinator of this project and in line with the obligations as stated in Article II.2.3 of the Grant Agreement declare that:

- The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;
- The project (tick as appropriate):
  - ☒ has fully achieved its objectives and technical goals for the period;
  - ☐ has achieved most of its objectives and technical goals for the period with relatively minor deviations[1];
  - ☐ has failed to achieve critical objectives and/or is not at all on schedule[2].
- The public website is up to date, if applicable.
- To my best knowledge, the financial statements which are being submitted as part of this report are in line with the actual work carried out and are consistent with the report on the resources used for the project (section 6) and if applicable with the certificate on financial statement.
- All beneficiaries, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SMEs, have declared to have verified their legal status. Any changes have been reported under section 5 (Project Management) in accordance with Article II.3.f of the Grant Agreement.

Name of scientific representative of the Coordinator: Silvio Paolo Sabatini

Date: 30th April 2010

Signature of scientific representative of the Coordinator: .....

[1] If either of these boxes is ticked, the report should reflect these and any remedial actions taken.

[2] If either of these boxes is ticked, the report should reflect these and any remedial actions taken.

1. Publishable summary

EYESHOTS is a Collaborative Project funded by European Commission through its Cognitive Systems, Interaction, Robotics Unit (E5) under the Information and Communication Technologies component of the Seventh Framework Programme (FP7). The project was launched on the 1st of March 2008 and will run for a total of 36 months. The consortium is composed of 7 research units of 5 research centres:

University of Genoa, Italy	(UG)
Westfälische Wilhems-University Münster, Germany	(WWU)
University of Bologna, Italy	(UNIBO)
University Jaume I, Castellòn, Spain	(UJI)
Katholieke Universiteit Leuven, Belgium	(K.U.Leuven)

which provide different expertise ranging from robotics, computer vision, neuroscience and experimental psychology.

1.1 Project's goal

The goal of EYESHOTS is to investigate the interplay existing between vision and motion control, and to study how to exploit this interaction to achieve a knowledge of the surrounding environment that allows a robot to act properly. Robot perception can be flexibly integrated with its own actions and the understanding of planned actions of humans in a shared workspace. The research relies upon the assumption that a complete and operative cognition of visual space can be achieved only through active exploration of it: the natural effectors of this cognition are the eyes and the arms (see Fig.1).

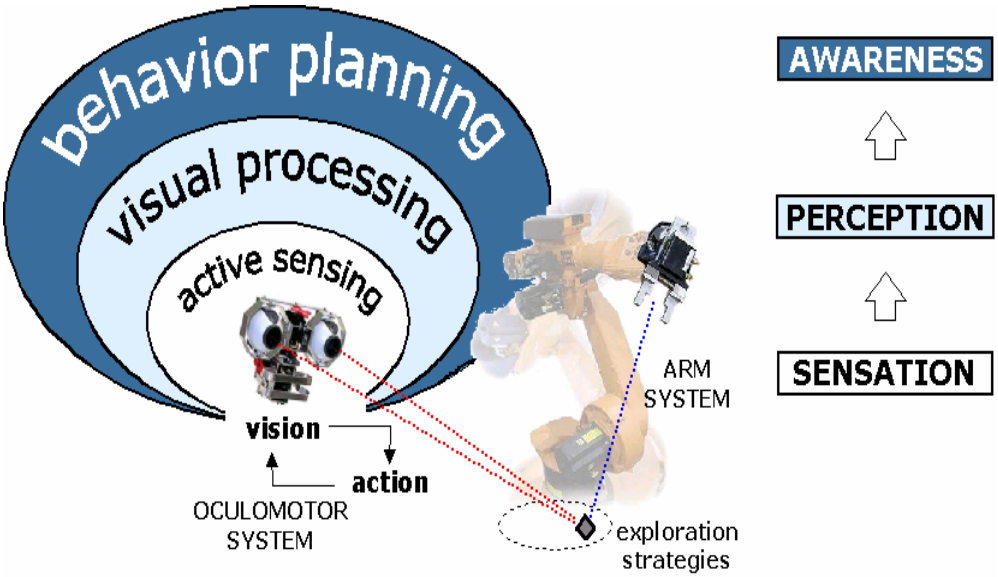


Figure 1: *The EYESHOTS perspective for 3D space perception.*

Crucial but yet unsolved addressed issues are object recognition, dynamic shifts of attention, 3D space perception including eye and arm movements, and action selection in unstructured environments. The project proposes a flexible solution based on the concept of visual fragments, which avoids a central representation of the environment and rather uses specialized components that interact with each other and tune themselves on the task at hand.
In addition to a high standard in engineering solutions the development and application of novel learning rules enable the system to acquire the necessary information directly from the environment.

1.2 Specific objectives

The project aims to reach the following three specific objectives:

Objective 1: Development of a robotic system for interactive visual stereopsis. The function of the systems is to interactively explore the 3D space by active foveations. Benefits of the motor side of depth vision are expected to be bi-directional by learning optimal sensorimotor interactions.
Objective 2: Development of a model of a multisensory egocentric representation of the 3D space. The representation is constructed on (1) binocular visual cues, (2) signals from the oculomotor systems, (3) signals about reaching movements performed by the arm. Egocentric representations require regular updating as the robot changes its fixation point. Rather than continuously updating based on motor cues or a visual mechanism (i.e. optic flow), the model updates only the egocentric relationship and object-to-object relationships of those objects currently in the field of view. During motion, the model covertly and overtly shifts attention to objects in the environment to maintain the model's current awareness of the environment. The updating of the internal representation of spatial relations requires binding processes across the different visual fragments.
Objective 3: Development of a model of human-robot cooperative actions in a shared workspace. By the mechanism of shared attention the robot will be able to track a human partner's overt attention and predict and react to the partner's actions. This will be extremely helpful in cooperative interactions between the robot and a human.

1.3 Expected final results

By the end of the three years the following results will be achieved:

- Implementing strong "dynamic" and "pro-active" components in which the effect of eye movements and of arm reaching actions will express as joint visuo-motor features, patterns and relationships for a perceptual awareness of space;
- Building a contingent knowledge of the sensorimotor laws that govern the relation between possible actions and the resulting changes in incoming visual information.
- Binding of objects into a global workspace for cognitive task control.

Although the project EYESHOTS has an explorative, pre-industrial character, the innovative computational paradigms and the cognitive engineering solutions, devised to operate adaptively outside the manufactured environments as well as pragmatic application scenarios, are expected to have impact on service robotics. From this perspective, we have been contacted by the international organization e-ISOTIS (Information Society Open To ImpairmentS, www.e-isotis.org), established and evolved with the scope to support the people with disabilities and elderly to overcome the existing barriers and have an independent living and quality of life, which is interested in the results of our project.

1.4 Work performed and main results achieved in the reporting period (01/03/09 – 28/02/10)

At the end of its first phase (February 2009), numbered among the project's assets were a front-end vision module providing a cortical-like representation of the binocular visual signal for both vergence control and depth estimation, and a conceptual framework for modelling ventral/dorsal interactions in reaching (and grasping) actions. Moreover, the experimental set-up for the planned (fixate-in-depth and reach-in-depth) neurophysiological experiments was defined, and a first set of psychophysical experiments on saccadic adaptation in humans was completed. At that stage, the eyes and the arm system were considered as separate effectors.

Starting from that ground, **in the second year** we more decisively addressed the problem of combining 3D space information obtained through active ocular and arm movements with the final objective of controlling spatially directed reaching actions, and, in general, visually-guided goal-directed movements in the whole peripersonal workspace. To this end, a first level of integration has been achieved both for (1) the *visuomotor coordination of eye movements* (K.U.Leuven, UG, WWU):

- Vergence and version behaviour to targets selected on the basis of object-based saliency criteria.
- Concurrent an iterative refinement of gaze and disparity estimates by using a variety of image warping mechanisms.
- Biomimetic kinematics of the eye movements and its relationships with visual perception.

and for (2) the *visual/oculomotor/arm-motor coordination* (UJI) by developing a network model with populations of radial basis function neurons uniformly distributed in the visual space (disparity and cyclopean position) as well as the vergence/version space and in the arm joint space.

Focusing on the action-oriented dorsal stream, neurophysiological findings from UNIBO, show that a large majority of cells in area V6A are modulated by ocular and/or reaching movements in 3D.
These results have contributed to the definition of joint representation signals of eyes and arm movements in a 3D extrinsic coordinate frame, on which to base the 3D location of a visual target with respect to the body. The advantages from such combined representation, respect to computing from signal intrinsic to each system such as version/vergence or joint-angle signals is currently under investigation.

Concerning the motor influence of visual perception, the analysis of the experimental data collected in the first year on saccadic adaptation is now completed. In the paradigm of saccadic adaptation a modification of motor parameters is evoked by the introduction of an artificial visual error after every saccade. The analysis have yielded remarkable evidences of the oculomotor components of visual target localization, which will be included in the model in the next period.

The actual stage of interaction of the different components into an architectural working model is summarized in Fig. 2. The different colors refer to two mutually interacting major complexes:

- the ***vergence-version control strategies with attention effect*** to guide active binocular exploration of the environment (work by K.U.Leuven, UG, WWU)
- the ***multimodal representation of space through concurrent reach/gaze actions*** (work by UJI, UNIBO, WWU).

The processing proceeds through a hierarchy of distributed representations in which sensorial and motor aspects coexist explicitly or implicitly. More specifically, the diagram evidences (thick-line boxes) a hierarchy of learning stages at different levels of abstraction, ranging from the coordination of binocular eye movements (e.g., learning disparity-vergence servos), to the definition of contingent saliency maps (e.g., learning of object detection properties), up to the development of the sensorimotor representation for bidirectional eye-arm coordination. The long-dashed arrows indicate motor influences on the distributed representations.
On our opinion, this can be considered, **from a methodological point of view**, an interesting intermediate result of the EYESHOTS project. Through the distributed coding, indeed, it is possible to avoid a sequentialization of sensorial and motor processes, that is certainly desirable for the development of cognitive abilities at a pre-interpretative (i.e., sub-symbolic) level, e.g., when a system must learn (binocular) eye coordination, handling the inaccuracies of the motor system, and calibrate the active measurements of the space around it.
From this perspective, early development of a sensorimotor intelligence is formalized and tackled as a problem that jointly involves three concurrent aspects:
Signal processing: by defining the proper descriptive elements of the visual signal (in the Gibsonian sense) and the operators to measure them (cf. the Plenoptic Function (Adelson & Bergen, 1991).
Geometry of the system and its kinematics: that directly relates to the embodiment concept. The measurement processes on the visual signal can be parameterized by specific geometric and kinematic conditions.
Connectionism paradigms: that defines neuromorphic architectural solutions for information processing and representation.
The connectionism paradigm (i.e., hierarchical, distributed computing) is crucial to guarantee accessibility and interaction of the information at different levels of coding and decoding, by postponing decisions as much as possible.

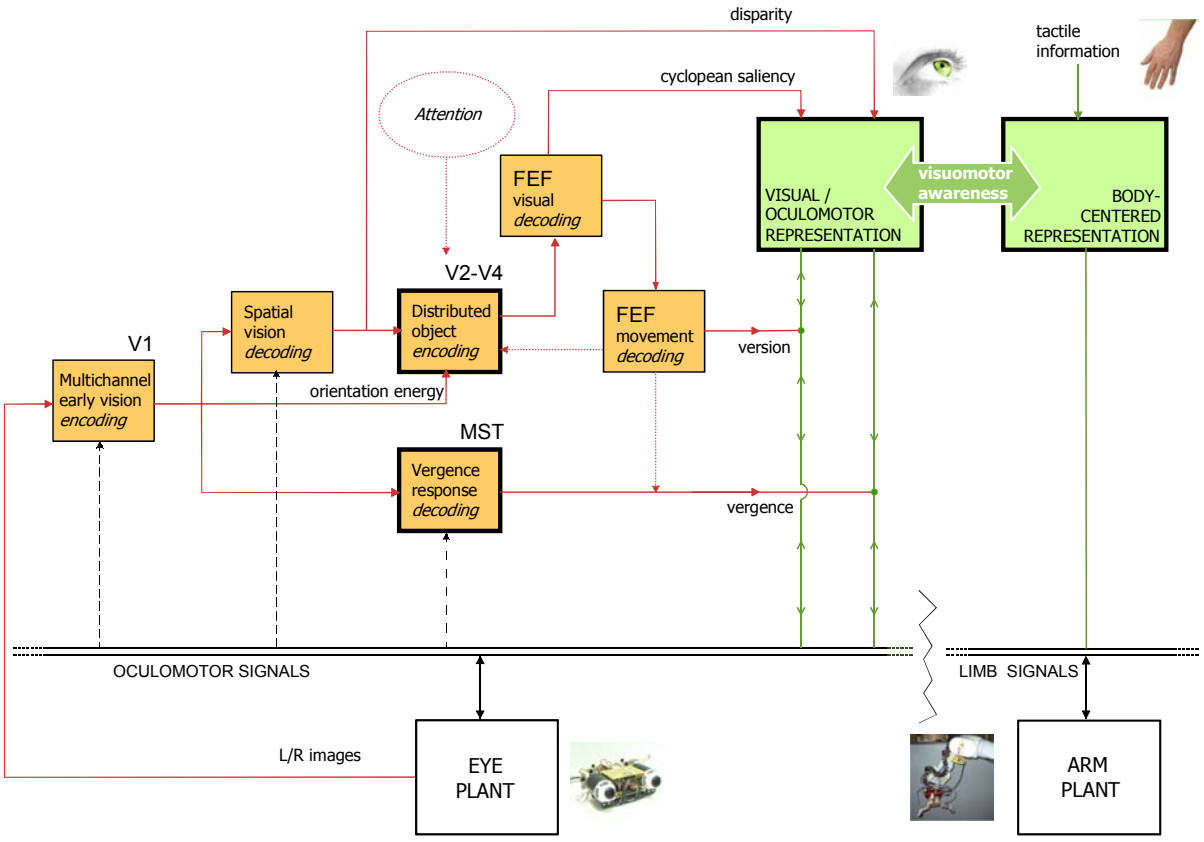


Figure 2. *Schematic diagram of the actual stage of integration of the proposed visuomotor architecture. Two cognitive modules (represented as the differently colored architectural complexes) mutually interact through the eye/arm end-effectors. The red architectural complex describes, at functional level, the developed modules for interactive visual stereopsis, and directly relates to project's Objective 1. The green architectutal complex, describes, at a higher level of abstraction the module subtending the ego-centric visuo-spatial awareness of the peripersonal reaching space, which directly relates to project's Objective 2.*

Among the further achievements of the second period:

- A first version of a model of Basal Ganglia has been developed and tested. In the next period the model is expected to be integrated in the overall architecture to allow learning of visual targets according to the task at hand by using unspecific reward, only.

- Data on human-human cooperative actions in a shared workspace have been collected and analyzed. The results showed that gaze and hand monitoring of a cooperation partner provide predictive cues that influence both quantitatively and qualitatively the interactive behaviour. Integration and testing of these non-verbal communication between human and robot setup is planned in the next period.

In summary, the project work during the two years went on schedule with the concerted work of all partners. The up-to-date results on the project work are available on the project web-site www.eyeshots.it

2 Project objectives for the period

2.1 Overview

The Work Program consists of 8 Work-packages (WPs). There will be five scientific and technological WPs (WP1-5), and three WPs, planned for: training (WP8), dissemination and exploitation of the project's results (WP7), and for general project coordination and management (WP6).

The workplan is organized to allow the *concurrent* development of these activities.
For each Workpackage we provide, from the Annex I of the GA, a synthesis of the objectives of the related tasks for the **2nd reporting period**.

WP1: Eye movements for exploration of the 3D space

(1) Study of the geometric and kinematic effects of eye movements on image flow for supporting the estimation of 3D information.

(2) Development of a bio-inspired stereoscopic robot system capable of emulating the ocular motions.

Task 1.2: *Perceptual influences of non-visual cues* – Direct and indirect perceptual consequences of eye movements. First, we will derive the relationships between the version and vergence angles and the locations of the fixated point in the 3D space, as well maps of binocular correspondences (horizontal and vertical retinal disparities) around the fixation point. Second, we will introduce specific mechanisms (e.g., gain fields) to modulate the responses of the disparity detectors of the visual system on the basis of the vergence and version signals.
Task 1.3: *Control of voluntary eye movements* – Study of the interplay existing between the mechanics of the eye plant and the strategies implemented by the brain to drive typical biological ocular motions. The goal is to understand the control mechanisms adopted by the brain to coordinate the action of the extra-ocular muscles for the single eye and for conjugate ocular movements.
Task 1.4: *Bioinspired stereovision robot system* – Design of a human-sized and bio-inspired binocular robot system capable of emulating basic ocular movements. The goal is to validate experimentally the models investigated in Task 1.3.

WP2: Active stereopsis

(1) Specialization of disparity detectors at different levels in a hierarchical network to see the effect of learning for the extraction of the binocular features.
(2) To learn a vergence motor strategy that optimizes the quality and efficiency of the feature-extraction for the specific task to be accomplished.
(3) To develop scanning strategies that accurately describe the head-centric disparity of a visual fragment so that it can be processed by precise, near-tuned disparity detectors.

Task 2.1: *Network paradigm for intelligent vergence control (reflex-like)* – To develop a convolutional network-based vergence control from a population of disparity-based detectors. The network is trained off-line, using a set of binocular images obtained from the VR simulators developed in the first year by UG and K.U.Leuven. In this phase, eye movements are simulated.
Task 2.2: *Interactive depth perception* – To develop a mechanism that renders the disparity-to-depth transformation robust to the limited accuracy of the motor system. Specifically, the mapping from disparity and eye position to 3D depth will be made insensitive to the effects of small vergence/version errors.

WP3: Selecting and binding visual fragments

(1) To derive information about object identity from a hierarchical representation of learned features.
(2) To learn distributed representations that actively bind and represent visual fragments for the task at hand. Reward-based learning approaches will be adopted.

Task 3.1: *Defining visual fragment: object identity* – Development of learning algorithms that allow us to encode object properties. The learning process starts from models of early visual areas, but it will then be applied at different levels of a hierarchical network to develop tuned receptive fields with increasing selectivity to complex features and disparity cues.
Task 3.2: *Selecting visual fragment* – Development of dynamic goal-directed attentional selection to bind object properties, on the basis of the momentarily existing task.
Task 3.3: *Selecting between behavioral alternatives* – We address the problem of learning the cognitive control of visual perception, in forms of visual-visual and visual-reward associations.

WP4: Sensorimotor integration

(1) Generation of an action-perception integrated representation of objects in the peripersonal space through the interaction of the robotic system with the environment.
(2) Achievement of an egocentric 3D visuomotor map of the peripersonal space to demonstrate binding capabilities in reaching and visual spaces.

Task 4.2: *Generating visuo-motor descriptors of reachable objects* – Describe a perceptual framework for the purposeful building of the representation of Task 4.1 through active exploration of the peripersonal space. Visuomotor descriptors based on modelling of cortical functions, mainly from the parietal cortex.
Task 4.3: *Constructing a global awareness of the peripersonal space* – The agent will simultaneously learn to reach towards different visual targets, achieve binding capabilities through active exploration and build an egocentric, 3D visuomotor map of the environment.

WP5: Human behaviour and neural correlates of multisensory 3D representation

Definition and execution of specifically-designed psychophysical and neurophysiological experiments. The experiments are intended to provide architectural guidelines for the organization of perceptual interactions and will guide the production of artificial systems able to explore and interact with the 3D world. The psychophysical experiments will provide behavioral patterns (I/O specifications), while in-vivo experiments will provide architectural solutions (I/O + internal structural data).

Task 5.1: *Role of visual and oculomotor cues in the perception of 3D space* –To verify the role of non-visual and visual cues in the perception of the 3D peripersonal space in the medial parieto-occipital cortex.
Task 5.2: *Link across fragments* – To experimentally determine neural correlates of multisensory representation of 3D space obtained through active ocular and arm movements.
Task 5.3: *Motor description of fragment location* – To experimentally determine motor contributions of eye movements to fragment location via saccade adaptation.
Task 5.4: *Predicting behavior and cooperation in shared workspace* – To study specific aspects of human behavior in the combination of allocation of attention and direction of gaze that can be used for prediction in human-robot interaction.

WP6: Project coordination and management

To implement and maintain an effective administrative and management infrastructure of the project, including:

(1) Continuous maintenance and update of the EYESHOTS project web-site (both the public section and the private section with restricted access to the consortium's members).
(2) Continuous maintenance of e-Services and repositories for broadcasting and sharing documents and data.

WP7: Knowledge management, dissemination and use, synergies with other projects

To make the project results known to the Community of interested researchers and automation industry as one of the potential developers of next-generation robotic systems.

Task 7.1: Regular publication of research news, events, research results, and demos on the project website.
Task 7.2: Organization of brainstorms events – as open workshop sessions – where, in particular, young researchers are invited to present their current work related to the project.

Task 7.3: Journal publications, participation to workshops, conferences, and other forum and events. Managing the mailing-list to disseminate results to interested parties.
Task 7.4: Synergies with other FP6 or FP7 projects.

WP8: Training, education and mobility

(1) To make local education, training activities and knowledge of the partners accessible for the entire consortium.
(2) To foster the exchange of personnel and to promote collaboration at every level of the consortium.

Task 8.1: To update the bibliography list and source/access information of the basic and recent literature relevant for the project.
Task 8.2: Student's half-yearly seminars.
Task 8.3: Medium- and long-term visiting periods by young researchers and short-term visits of principal investigators.
Task 8.4: Organization of the Summer School.

Summarizing, for what concerns the main S&T issues, the project's objectives for the 2nd reporting period were:

1. The concurrent development of models and architectures for the oculomotor control and interactive stereopsis, including early integration of motor components and attention signal (vergence/version and depth vision). Work performed by UG, K.U.Leuven, WWU.
2. Collection of experimental evidences about visuomotor integration mechanisms in primates and humans: (i) Neurophysiological evidences of 3D visuomotor integration in area V6A; (ii) Motor contribution of eye movements to fragment location via saccade adaptation.
3. Achievement of first results on a heterogeneous representation of the peripersonal space and definition of the experimental set-up.
4. Development of models of human-human interaction in shared workspace and definition of the human-robot interaction experimental set-up.
5. Start prototype realization of the mechatronic robot eye system.

For what concerns the other, management, issues the main project's objective was:

1. The organization of the Summer School.
2. Dissemination activity, e.g., through joint publications.

All of these objectives have been achieved.

Concerning the detail of the individual objectives, they are well documented in the individual workpackages sections and summarized in section 3.2.

2.2 Follow-up of previous review

There were two main recommendations and one minor comment, the Consortium was asked to take care of. These are reported here below in the greyed boxes from the first Technical Review Report.

Recommendation no. 1

Identification of an experimental task: The reviewers appreciate the statement made by the consortium that to keep the generality of the research, they would like to avoid over-specifying a task that may force them to make assumptions impinging on their ability to make general claims regarding their findings. However, a task even as general as possible would enable them to demonstrate their claims and the above mentioned constraints (no neck, etc) would then make sense in the context of such a well defined task. Whilst the reviewers would like to encourage them to think of a general task (possibly involving grasping as a "given" capability rather than as an additional research topic -- it is the reviewers' understanding that such a capability is available to UJI), their recommendation is that a suitable form of demonstration be that:

- the system exhibit fixating on and pointing to some visual stimulus (e.g., a new object in the environment or any object of a particular characteristic, e.g., the green object) AND
- this visuo-motor behaviour be implemented through mechanisms that are grounded in physiologically-plausible mechanisms, in particular, that they capture key processes regarding the bilateral interaction between motor and perceptual processes. For example, such a demonstration could be such that it can be repeated with the monkeys using LEDs set up on a panel. The consortium is invited to consider the best way to illustrate that the demonstrator captures the animal physiological mechanisms for example by considering some suitable form of visual adaptation paradigm (e.g., monovision, foveal/peripheral vision -- see point iii as well, prismatic adaptation, local deformation of the visual field) that will not be built in the system but in response to which it will be shown that the system exhibit the same response as that of the animal, i.e., it should be verifiable and testable across all domains.

NB1: Ideally, testable predictions will be made, but not necessarily tested within the lifetime of the project. It is important to stress that the contribution of the project cannot simply be evidence of having learned eye-hand co-ordination but instead the focus must be on the concept of fragments.

NB2: An additional consideration worth considering in defining the task should be the relationship between the two motor systems considered: the visuo-motor platform, and the arm. Whilst manipulating the object being fixated would be a natural way of looking at the problem, an alternative would be to consider the idea of a 'sticky hand' so that once the arm reaches an object, any subsequent movement would result in the fixated object smoothly moving in the visual space. Tracking of the object would be equivalent to vergence control over a smooth surface in virtue of the continuity of control.

NB3: The question of whether using a peripersonal space as part of the demonstration should be evaluated in terms of whether the benefits outweigh the added complexity.

The capability of planning reaching actions through fragmented vision (i.e., stability of the "reaching maps" by updating the egocentric relationships among the objects currently in the field of view) is one of the peculiar project's objectives. From this perspective, the relationship between the oculomotor system and the reaching system is fundamental for the project. Since the focus of the work is on saccadic movements, not continuous tracking, we have adapted the concept of the sticky hand to the saccade/reaching case. The core of this idea is that information from the hand (e.g. current hand location, current joint angle configuration) can be used to control the gaze movement. This is implemented in the model by the inverse transformation, which accepts a joint angle configuration and calculates the corresponding version/vergence configuration. By using this transform, the model can follow the hand location without using visual input by directly employing somatosensory or efference copy signals. A movement of the hand would thus result in a coordinated movement of the eyes to continue to follow the current hand position with a saccade or sequences of saccades. Also, gazing or reaching movements can be planned but not released, allowing for covert transformations and more complex behaviors, such as peripheral arm reaching movements. The basic visuomotor capabilities of the system are acquired by performing concurrent gaze/reach actions, and thus integrating different visual fragments associated to corresponding oculomotor configurations.
Taking all of the above into account, **we devised a number of increasingly complex and scientifically interesting tasks to be performed by the system to illustrate its sensorimotor skills.**

- As a first experimental task, the system can be required to show its visuomotor capabilities by performing an oculomotor action toward a visual target or toward the location where its arm lies. Complementarily, it should also be able to perform arm reaching movements to a visual target, either with or without gazing at it. In the latter case of peripheral reaching, an intermediate transformation from visual to oculomotor space is performed, but the corresponding motor signal is not released.
  The underlying computational model which allows the system to perform such tasks is built on a basis function representation able to transform between, and contextually code for, different reference frames. Neural populations are established on the base of electrophysiological findings regarding posterior parietal area V6A, and the initial setup resembles those employed in monkey studies. The model and its robotic implementation are expected to be able to reproduce some of the experimental protocols and related effects used in WP5. For what concerns monkey studies, the robot should be able at least to exhibit the skills described in Tasks 5.1 and 5.2, including vergence control to reach at different depths. Regarding psychophysical effects, the system is expected to show a simple form of saccadic adaptation, to illustrate the way altered/deprived perceptual conditions affect the functioning of the model and its adaptation skills. Analysis of how deceptive visual feedback (as in the saccadic adaptation protocol) will affect the artificial agent oculomotor and arm motor abilities will serve as a validation of the underlying model, and may help in advance hypothesis on saccadic adaptation mechanisms in humans and monkeys.
- The second experimental setup will see the inclusion of multiple objects within the peripersonal space of the system, which can be asked to perform any of the actions of the first setup to different goals, for example in a given sequence or on demand, or even to reach/gaze toward objects out of the field of view by using a sensorimotor memory of the environment. This step requires the integration of different perceptual fragments into a global spatial awareness of the peripersonal space.
- The third and final experimental setup will include human-robot interaction tasks, in which the actions performed in setup 2 toward different goals will be guided by human gaze direction and possibly also by arm movements. For example, the robot can be required to reach toward the object/goal position the human partner is gazing at, or to gaze and reach toward an object that has been just released by the human. Experiments on human-human interaction performed by WWU will be taken as inspiration for this final experimental step.

Recommendation no. 2

> **Increased integration between partners:** The project features research spanning neurophysiology, psychophysics, computational neuroscience, robotics. For the project's scientific objectives to be achieved, increased integration between partners is needed to guarantee that findings by each partner tie-in in a convincing manner. For example, there should be **closer synchronisation between WPs dealing with disparity filters (WP2 and WP3).** The crux of the reviewers' recommendation, however, is in regard to the **integration of UJI with the rest of the project.** Because of its central position in terms of putting together the different strands of the projects, and its stated objectives of defining biologically plausible integrated representations, care must be taken that design decisions (e.g., choice of modelling techniques, data analysis) must be matched against what is available to the experimental component of the project in terms of validation. How to best reconcile different disciplinary approaches? For example, the global architecture of WP4 combines multiple sub-areas of the visual cortex when the neurophysiological data relates to a number of neurons in a specific area. It is the reviewers' opinion that WWU should play a pivotal role in guiding, supporting and enhancing the interaction and transfer of knowledge between the neurophysiological component of UNIBO and the modelling component of UJI. Some PMs could be shifted to provide support for WWU and UJI to work more closely together. In particular, it is felt that having personnel (e.g., postdoc) from WWU visit UJI would be beneficial to preventing UJI from diverging. Such a divergence will be difficult to avoid because of the inherent granularity at which the respective partners work. It is thought that with WWU having expertise at both ends of the spectrum, they will be in the best position to provide sanity checks.

On the integration among WPs, concrete progresses have been done. Specifically, concrete measures have been undertaken in months 19-21 to define a proper modeling framework for an integrated representation of the 3D space. Major effort has been dedicated to this task: a new PhD student has been hired by UJI, a collaborative effort has been shared between UJI and WWU (Eris Chinellato, Fred Hamker and Markus Lappe). First results were achieved. In order to provide support for WWU and UJI to work more closely together, a minor number of months has been shuffled across WPs (in particular, UJI moved 2 PMs from

WP1 into WP5, similarly, WWU moved 3 PMs from WP5 into WP4). These shifts make the distribution of the person months better adapt to the current status of the workplan without affecting global equilibrium among partners. Concerning the synchronization between WP2 and WP3, an integrated architecture has been now developed (see also p. 24), which share the same early vision processing stages for disparity detection.

Recommendation no. 3

> Finally, on a minor note, it is recommended that the consortium explicitly indicates that they will be using a **non-linear optical system** (e.g., the simulated log-polar mapping available at UNIGE) as a way to provide a consistent level of complexity between motor platform and optical system. It is our understanding that this log-polar mapping is already in use in the team of the co-ordinator.

A space variant sensing of the visual signal is an intrinsic feature of any active vision system (cf. foveations). In the first phase of the project, space variant processing is assumed to be potential, and stays in the background, without affecting the general approach of the proposed models and architectures. In agreement with the reviewers' recommendation, the importance of an *explicit* use of the space-variant sensing has now been stressed. Accordingly, a front-end software module performing a log-polar retinocortical transformation has been specifically developed and tested by UG, and shared within the Consortium. The novel design rules adopted for the mapping and the filters guarantee the minimal distortions of the mapped filters across the retinal eccentricity and allow their direct use for feature extraction. The research activity on this topic has been conducted within WP2 and therein reported (see p. 52), moreover, a publication is currently under revision to Pattern Recognition Letters (Solari et al, 2010 [J8] The combined disparity and gaze estimation procedure proposed by K.U.Leuven in Task 2.2 already operates in the cortical domain on the basis of the same log-polar transformation used by UG. Steps towards the integration of the log-polar space-variant sensing in WP3 and WP4 already started. If necessary, possible diversification of the mappings to be adopted in the different levels of a general architectural model of cortical processing (cf. different cortical areas, in WP3 and WP4) will be considered.

3. Work progress and achievements during the period

Note: a "code" and a "number" label the publications of the Consortium in the reporting period: the codes 'P' and 'C' refer to journal paper and conference contribution, respectively. For the full list of publications please see section 5.2.

By updating the state of the art about some ***key concepts*** on which the project's research is based, we can state that: (1) the grounding concepts of EYESHOTS are still valid, and (2) there have not been any advances in the field that would force EYESHOTS Consortium to restructure the work. On the contrary, the project's objectives are even more actual and challenging.

A visuomotor approach to 3D perception

In absence of motion parallax, and disposing of a binocular stereo vision system, 3D information can be gathered without introducing active components (except for covering a larger field of view). In such conditions, stereopsis (i.e., position parallax) is usually thought as a static problem, since the disparity map obtained by a fixed-geometry stereo camera pair, with (nearly) parallel axes, is sufficient to reconstruct the 3D spatial layout of the observed scene, and eye movements are often regarded as unnecessary complicating factor. Things are dramatically different if we consider a binocular foveated system with a vergent stereo geometry. In such conditions, the perceptual process intrinsically gains a motor dimension as the 3D information is collected dynamically with respect to the fixation point. The eye rotations, although insufficient *per se* to provide depth cues (but see Santini & Rucci, 2007), strongly constraint the stereo vision processing and visual information is gathered through a continuous interaction with the environment. This is especially true for visual exploration of the peripersonal space when large values of vergence occur.

In general, the importance of the visuomotor aspects of 3D visual perception can be understood not just to develop algorithms that enable robots to cope with changes in the environment, but and more importantly, to acquire task-independent skills as a living being. Indeed, if the problem of controlling on a visual basis the vergence of a stereo camera system has specific and rather straightforward solutions, the joint treatment of the vergence control and of 3D perception still represents a challenging cognitive problem. The zero-disparity condition in the fixation point solves indeed the vergence task, but nullifies the visually-based information for the 3D position of the fixated target point. Only the residual disparities elsewhere in the visual field are cues for stereopsis. The momentarily existing (and continuously changing) fixation point, i.e. *where* the system verges, becomes a reference that can be parameterised by the relative orientations of the eyes. Moreover, *how* the system verge has an impact for the accuracy of stereopsis. Different camera positions (and mainly torsional postures) influence the local shape of the zero disparity surface near the fixation point (i.e., the surface horopter) with several expected advantages on vision, such as an optimal use of the range over which the disparity detectors operate, or a "corrective" warping of the images to adjust the slant of the observed surface when it deviates from the stereotypical frontoparallel case. Although these facts have been understood in the psychological literature for over a long time (Howard & Rogers, 2002), it is only recently that computational and theoretical approaches attempted an engineering formulation of these concepts in order to provide operative guidelines to quantitatively analyze their visual (and motor) advantages, and optimal design criteria for active artificial vision systems (Schreiber et al., 2001; Schor et al., 2002; Read & Cumming, 2004; Hansard & Horaud, 2008; Hansard & Horaud, 2010). From a behavioural point of view, the empirical derivation of the horopter has been used to simulate the optimally stereo-viewed surface in foveal vision (Schreiber et al., 2006; Schreiber et al., 2008), and complementary non-visual information exploited to estimate absolute distances when the system is engaged in reaching tasks towards non foveally viewed targets (Greenvald & Knill, 2009; Blohm et al., 2008). From a computational point of view, a debate on the role of vertical disparities to calibrate depth perception pointed out that the vertical disparity is not simply tolerated, but it is actively detected and used in perception (Read & Cumming, 2006; Read et al., 2009; Serrano-Pedraza & Read, 2009; Serrano-Pedraza et al., 2010) and to guide vergence behaviour (Yang et al., 2003; Sheliga & Miles, 2003; Gibaldi et al., 2010 [C21], but see Rambold & Miles, 2008). As a whole, the 2D vector disparity pattern should be recovered when incorporating eye movements since limiting the search of the binocular correspondences on the epipolar lines predicted by the current position of the eyes would be vulnerable to any inaccuracies in the sensed eye position. Exploiting mutual dependencies between the disparity patterns and the epipolar geometry e.g., by adapting the computational

resources or the processing to the fixation constraint, can be a viable solution to the inaccuracy of eye position sensing (cf. the approach proposed by K.U.Leuven and UG in WP2, see Chessa et al., 2009c [C6] and p. 47 in this report). In general, to overcome the difficulties in obtaining an accurate estimation of epipolar geometry, Monaco et al. (2009) demonstrated a symbiotic relation between foveation and uncalibrated active vision systems to minimize the number of points per epipolar space and thus improve the efficiency of the search for stereo matches.
Altogether, there is a ample evidence for the pivotal role of programmed eye-movements in the computations that are performed in the process of seeing (as opposite to "looking at"). Yet, how to profitably integrate these accumulating evidences with the computational theories of stereo vision has not been fully exploited, as rectified images are still calculated in current humanoid active-disparity vision modules, relying on the encoders data, only (see e.g., iCub Software Repository Documentation[3], 2010). The complexity of integrating efficiently and with a proper flexibility the different aspects related to a binocular active behaviour prevented till now a full validation of the visuomotor approaches to 3D perception in real world situations.

Motor components of visuospatial awareness in the peripersonal space

- **The role of eye movements (i.e., the "active fragment paradigm")**

When an observer scans a visual scene a new partial image is typically acquired every 200-300 ms by a saccadic eye movement which directs gaze to a part of the scene that is anticipated to provide additional information for the scene representation (Henderson, 2003; Hayhoe and Ballard, 2005). During one fixation, scene perception can be very fast (Rousselet et al., 2004) if humans know in advance what to look for. However, inattentional blindness experiments have shown that humans can miss even substantial aspects of a visual scene, when engaged in a demanding task (Most et al., 2005). These findings demonstrate that the subjective richness of the visual representation of the world at any time is an illusion. Perception is not only an active act in which we continuously sample the scene by rapid eye movements, but it is also an anticipatory, cognitive process in which generated expectations are related to the visual input. Until recently it was believed that the brain constructs a composite image of the external world by combining images from consecutive fixations (Jonides et al., 1982), but little experimental evidence supported this idea. For example, subjects do not fuse visual patterns (O'Regan and Levy-Schoen, 1983) and do rarely notice scene changes across saccades (Henderson and Hollingworth, 2003). Thus, it seems that every fixation provides a new piece of local information, called a visual fragment, about a certain part of the scene, and that these fragments stay to some degree separate across saccades.
While general scene changes are often unnoticed during a saccade, changes at the saccade target are noticed and pre-view experiments suggest that some aspects of the current view of the scene (the current fragment) influence the perception during the subsequent fixation. Most fragile during a saccade is location information. Intra-saccadic displacements of visual stimuli, even of the saccade target, are usually not noticed. When the saccade target is repeatedly displaced, however, humans unconsciously adapt the gain of the saccade and compensate for the visual expectation that the postsaccadic saccade target should appear in the center of the visual field at the new point of fixation (McLaughlin, 1967). This adaptation of oculomotor parameters recursively modulates perceived location of objects in the scene (Bahcall and Kowler, 1999; Awater et al., 2005), which suggests the existence of a closely knit integration of action and cognition in scene perception.

Research within the consortium and by other labs has, over the last years, provided more support for this close interaction and advanced the view that motoric information from saccade planning is part of perceptual localization (Lappe, 2009). Collins et al (2007) showed that localization of targets at various places in the visual field is affected by adaptation of a single saccade in proportion to the amount of transfer of the adaptation to saccades towards the respective target locations. The contribution of oculomotor information on localization is specific to the type of saccade (whether reacting to suddenly appearing targets, or scanning a stationary scene) and the visual properties of the localization object

[3] http://eris.liralab.it/iCub/dox/html/group__icub__DisparityMapModule.html

(flashed or stationary) (Zimmermann & Lappe, 2009 [J11]). Adaptive increase of saccadic amplitudes transfers to anti-saccade generation (Panouilleres et al., 2009), to hand pointing (Hernandez et al., 2008), and to perceptual localization even during fixation (Zimmermann and Lappe, 2010 [J12]). All of these studies underline how perceptual space is linked to motor knowledge as has been proposed by sensorimotor interaction theories of perceptual experience (O'Regan & Noe, 2001; Varela et al. 1992).

- **The role of arm movements**

The dimensionality of the external space is defined with respect to anthropometrical and motor factors (Coello, 2005; Coello et al., 2008), which postulate interaction with the environment for an incremental construction of the 3D spatial representation. Specifically, the peripersonal space is intended as the space where objects can be grasped and manipulated, i.e. the space within reach of any limb of an individual. Although unitary when examined introspectively the representation of space, it is discrete and fragmentary as based on numerous spatial maps that control different movements, such as eye movements, head movements, and arm movements. By locating sensorial information in a head or body-centered frame we develop the awareness about the distinctive and stable properties of the environment as well as the motor/cognitive skills to interact with. The link between visual and motor information lies in the fact that when we look at the target, we immediately gain its motoric significance: it evokes all the potential actions that we can make towards or through it. In this way, the observed objects cease to exist independently of the observer but has relevance only with respect to their interactivity. Points in space will be thus no longer considered as a geometrical entities but as a *goals* of movement actions (Rizzolatti et al., 2006). Current realizations of reaching and grasping behaviors on humanoid robots mostly rely on a-priori knowledge about the system (at least, a kinematic model of the robot) and about the environment (reconstructed 3D model of the scene, a-priori knowledge of the object properties) (Saxena et al. 2008; Rasolzadeh et al. 2010). More recently, in the field of cognitive robotics, a great effort has been devoted to develop developmental strategies to learn a kinematic model of the system, as well as other non-linear sensorimotor mappings, This learning has been performed either off-line using previously collected sensory measurements (Natale et al., 2007) or on-line during a training phase, separated from the subsequent execution phase (Rougeaux & Kuniyoshi, 1998). Applying on-line learning and adaptation, become crucial when we deal with dynamically changing environments. Considering humanoid robots, efficient reaching can be a basis for developing higher order cognitive processes (e.g. classify and label grasped objects (Jamone et al., 2006), learn about objects dynamics by tapping them (Fitzpatrick et al., 2003)) and social behaviors (e.g. by getting in physical contact with humans, by manipulating the same common-use objects). From this perspective, the incremental building of a reachable space map based on motor information, which would allow the robot to evaluate its possibility to reach for an object by fixating it is a crucial issue.

Summarizing, The understanding of the tight inter-relationships between action and perception is expected to improve the visuospatial perceptual capabilities of artificial systems according to a form of embodied cognition. Specifically, the methodologies for cognitive development in robots are used to overcome current limitations in robot design. To advance our understanding of cognitive development, a popular approach (cf. recent FP6 and FP7 projects in Cognitive Robotics: RobotCub, ITALK, IMCLEVER, GRASP …) proposes the study of artificial embodied agents (e.g. either robots, or simulated robotic agents) able to acquire complex behavioral, cognitive, and linguistic/communicative skills through individual and social learning. Specifically, to investigate sensorimotor (perception/action) integration, it is possible to design cognitive robotic agents capable of learning how to handle and manipulate objects and tools autonomously, to cooperate and communicate with other robots and humans, and to adapt their abilities to changing internal, environmental, and social conditions.

In the second year, the Project work has been focussed on:

- the rooting of these key concepts in the architectural components of the models
- a first integration of such components at system level
- the collection of experimental evidences to be included in the models.

3.1 Progress overview and contribution to the research field

With reference to the overall architecture of the EYESHOTS' agent, Fig. 3 depicts the status of the project. The research activities that have characterized the second reporting period are highlighted in yellow and they can be grouped in four parts:

- Part 1: 3D vision by interactive stereopsis (WP1, WP2, WP3)
- Part 2: Perception across visual fragments ["active fragmented vision"] (WP2, WP3)
- Part 3: Integrated sensorimotor representation of the 3D reaching space (WP4, WP5)
- Part 4: Cooperative actions in a shared workspace (WP4, WP5).

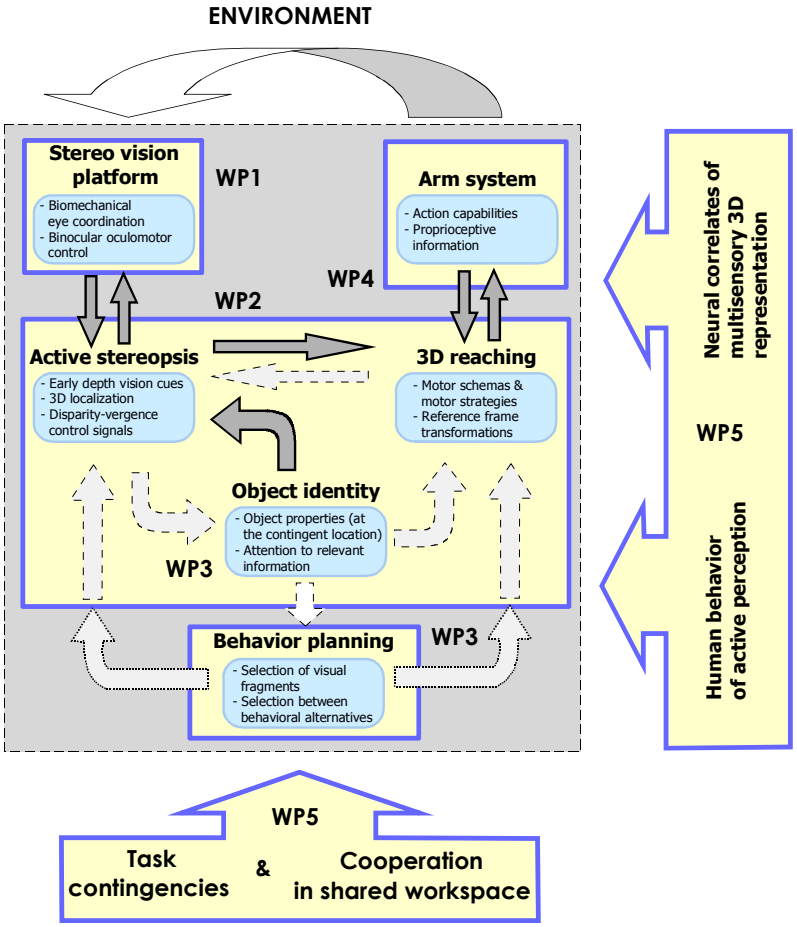


Figure 3: *Status of the project at the end of the 2nd period. The research activities that have characterized the period are highlighted.*

The progress of work for each part can be summarized as follows.

Part 1: 3D vision by interactive stereopsis

- ***Visuo-Motor constraints in binocular eye coordination***

  Partner UG proposed a mathematical framework to investigate the functional implications of the rotations of the eyes described by the Listing's Law (LL) for binocular vision.

  Starting from the seminal work of Tweed (Tweed, 1997), UG revisited the visuo-motor theory of optimal binocular control by generalizing the visual constraint to include the coplanarity of the fixation planes. Contrary to starting from assumptions on the postures of the eyes (Tweed, 1997) and analyzing their perceptual implications (e.g., see Read & Cumming, 2004), we proposed general design criteria by which both vision and motor efficiency principles would guide proper eyes' postures. In perspective, this approach would allow us to better take into account the resources that motor and vision systems have at disposition for the design of an artificial system.

  In particular, UG found, as a result, (1) the identity of Helmholtz torsions (T_l=T_r), postulated by Tweed (Tweed, 1997), for different instances of the visual constraint; and (2) the proportionality relationship, between the rotation of the Listing's planes and the vergence angle.

  Specifically, the effects of the different components of the cost functional are quantitatively evaluated with respect to (1) the relationships between the temporal rotations of the Listing's planes and the vergence and version angles, and (2) the amount of the resulting rotation eccentricities of the eyes. The major contributions of the work are: (1) the derivation of a general expression of the orientations of the eyes, which is dependent on the coordinates of the fixation point only, and not on the adopted rotation system; (2) the solution of the visuo-motor optimization functional directly with respect to the rotation angles of Listing's planes, and (3) the derivation of a computational justification of the compromise between LL and its binocular extension (L2) that takes into account most of the experimental evidences on coordinated binocular eye movements.

  Comparative assessment with the current state of the art: From the first time Tweed proposed the visual-motor theory, this has been taken as reference to explain L2. In reality, the same Tweed and many other researchers (Tweed, 1997; Schreiber et al, 2001; Schreiber et al., 2006) was always of the opinion that eyes do not follow L2 precisely, but they strike a compromise between the motor advantages of LL and the visual advantages of L2. Indeed, we have verified that it is not convenient to deal with a monocular LL, valid for null vergence, and a binocular version of it (L2), valid for near fixation. We argue that only a "generalized" Listing's Law (LLx) can be considered, forming a continuum that ranges from large to small vergences, and that includes in itself all the observed experimental evidences. It is not surprising that in our argumentation the geometrical behavior of a standard tilt-pan system results as the extreme case of the extended visual constraint. Though, it is worth underlining that such a configuration links up with a Listing-compliant one when the value of the vergence angle decreases. The way by which one configuration turns into the other depends on the relative importance/weight of the 'motor' and 'visual' constraints. From this perspective, the analysis justifies that it might be functionally useful to dispose of a robotic platform capable of mimicking human-like binocular eye movements, since that would guarantee an optimal flexibility, being the *task* to condition the behavior of eyes' movements.

  As a possible alternative to mechanical implementation of Listing's law, Miles and Horaud (Hansard & Horaud, 2010) recently proposed a systematic formulation to synthesize its effect on a set of calibrated images obtained from a standard (tilt-pan) robotic camera mounting. To this purpose, the torsion angle function, depending on the visual direction, has been derived and then the associated rotation matrices for a full rotation of the eyes. The principles and the relations derived by UG for binocular near vision, extend and complement the treatment of (Hansard & Horaud, 2010) by including the vergence angle, and can be directly used in their approach for further analysis of human-like binocular kinematics in robot heads.

- ***Disparity and gaze estimation***

  The problem of disparity estimation from uncalibrated or imprecisely calibrated stereo rigs has received considerable attention recently in the computer vision domain (see e.g. Dang et al., 2009; Wan & Zhou, 2010). All the proposed methods employ a strict separation between calibration-

correction and disparity-estimation. This is on one hand due to the application of well-understood procedures for geometry estimation from sparse correspondences, and on the other hand due to the discrete optimization techniques commonly used in state-of-the-art dense stereo algorithms, which are too computationally expensive to estimate vector disparity and thus require rectified image pairs. The phase-difference based technique used in our methods does not suffer from this shortcoming and allows us to solve both problems simultaneously. In addition, the proposed methods in Task 2.2 can operate directly in the cortical domain, which increases the efficiency and calibration errors the method can cope with.

- ***Progress on learning adaptive vergence behavior***
  Most of the conventional vergence control models (Westheimer & Mitchell, 1956; Rashbass & Westheimer, 1961; Krishnan & Stark, 1977; Schor, 1979; Hung, *et al.,*1986; Pobuda & Erkelens, 1993; Theimer & Mallot, 1994; Patel *et al.,* 1996; Horng *et al.,*1998), are based on the minimization of the horizontal disparity. We propose to avoid the explicit computation of the disparity map and extract the desired vergence angle directly from the distributed disparity representation. Two neural models for vergence control have been proposed and tested. Both models work in a closed-loop manner and use distributed disparity encoded in a population response of binocular complex cells (Chessa *et al.,* 2009a [C1]) for the vergence control. The actual gaze direction and actual vergence angle are used as the additional inputs to the vergence control network (these inputs will be dropped in the dynamical version of the vergence control). Experiments with different types of stimuli as well as the different patterns of stimulus motion (in depth) were carried out in order to compare the proposed approaches and assess their performance. Although the paradigm only resorts to a population of neurons in a single scale, we demonstrate that, using a neural networks, accurate and fast vergence control can be achieved in a closed loop, for different orientations of the gaze.

  Comparative assessment with the current state of the art: In a framework of reinforcement learning a reward mechanism the learning of disparity-based vergence control is proposed in (Franz & Triesch, 2007). A multi-layered network with feedforward and feedback connections learns to perform vergence eye movements between natural images of objects by receiving a reward whenever an object is fixated with both eyes. Disparity tuned neurons emerge robustly in the hidden layer during learning. Even though the proposed reward mechanism is biologically plausible, it develops the vergence control which relies on the neurons tuned to zero disparity only. The method works with one-dimensional (horizontal) slices of the input images, which limits its application to real-world active vision systems. The vergence control is achieved in terms vergence movements of fixed step (–1.0°, –0.8°, –0.6°,…, +1.0°) only, which is also a limitation of the proposed method.

  In (Wang & Shi, 2009) a reinforcement learning approach is used to for the training of a disparity population-based vergence control model. By applying a policy that maximizes the total response across the neuron population, the system eventually develops ability to track a moving (in depth) object. Even though the proposed approach shows on-line adaptation and is biologically plausible, nevertheless it has some flaws: 1) the stereo setup used in the paper is not active: the cameras are fixed and the vergence is emulated by moving subwindows in the rectified left/right images. This simplification eliminates vertical disparity, which might influence vergence; 2) the subpixel vergence is not discussed, which given the relatively low resolution (648x488) of the camera can lead to inaccurate vergence; 3) the system is trained to verge on a specially designed stimulus (a vertical dark bar on a white background), no other stimuli discussed; 4) the vergence is shown only for stimulus moving in depth sinusoidally with fixed frequency and amplitude, no other patterns of stimulus motion are presented.

Part 2: Perception across visual fragments ["active fragmented vision"]

- ***Visual feature extraction in foveated systems***

  During the second year, partner UG has addressed the problem of the multi-orientation and multi-scale filtering in the log-polar domain. To this aim, a systematic analysis of the relationships between the parameters of the discrete log-polar mapping and of a bank of Gabor filters has been carried out. The major outcome of this analysis is the definition of a set of general design rules, that allow us to use algorithms, which were originally designed in the Cartesian domain, directly in the log-polar space, without requiring specific modifications. Moreover, we have deduced a novel rule to efficiently implement a multi-scale analysis, by exploiting the space-variance of the log-polar mapping. The validity of such analysis has been proved by applying the distributed phase-based approach for the computation of vector binocular disparity based on a bank of Gabor filters (Chessa et al., 2009a [C1]) on log-polar stereo pairs. The obtained results show that it is possible to recover reliable values of the horizontal and vertical disparities by directly applying the algorithm in the cortical domain, thus achieving a consistent reduction in the execution time. The possibility of efficiently exploiting a space-variant representation is of great importance in the development of active systems capable of interacting with the environment, since a precise processing of the visual signal is possible in the foveal area, where the feature errors are small enough to allow a fine exploration of the object of interest. At the same time, the coarse computation of the feature in the peripheral area provides enough information to detect new saliencies and to bring the focus of attention there. The results are currently under review (Solari et al, 2010 [J8]).

  Comparative assessment with the current state of the art: The log-polar imaging is a well established paradigm for simplifying a wide number of computational problems in pattern recognition and active vision (Berton et al, 2006; Traver & Bernardino, 2010). The log-polar mapping simultaneously provides a wide field-of-view, high spatial resolution on the region of interest, and a significant data reduction. All these features are well suitable for active vision applications, since the visual systems continuously interact with the environment, by purposefully moving the eyes, to bring the interesting objects into the foveas. In the literature, many approaches to directly solve image processing and image understanding tasks for space-variant representation of the visual signal have been described (Fischl et al., 1997; Traver & Pla, 2008). Moreover, several authors in the literature addressed the issue of disparity estimation in the log-polar domain (Manzotti et al, 2001; Schindler, 2006).

- ***Object-based top-down selection***

  Visual attention refers to our ability to focus on only a small part of the vast amount of information obtained from our visual environment. The neural core of this ability relies on gain control, that is, mechanisms that alter the input-output ratio of individual neurons in visual brain areas. Attending to a particular location in visual space enhances neural responses to stimuli presented at this location, known as spatial attention [for a review see Reynolds & Chelazzi (2004)]. Similar effects have been reported if attention is directed towards a specific, non-spatial attribute (Chelazzi et al., 1993; Treue & Martinez-Trujillo, 1999; Martinez-Trujillo & Treue, 2004; Bichot et al., 2005). This feature-based attention is supposed to operate globally throughout the whole visual field, that is, the activity of neurons is enhanced if their tuning characteristics match with the attended feature even if their receptive field location is distinct from the present focus of spatial attention.

  Recent models of feature-based attention have been shown to allow a selection of simple early features (Zirnsak & Hamker, in press, Hamker, 2007; 2006), although the principles generalize to more high-level properties of objects. These properties are presently not features of biologically motivated object recognition models (e.g. Serre et al., 2007).

- ***Towards a context-based (and task-contingent) behavior planning***

  One of the main current challenges of artificial intelligence is about providing an agent the ability to behave autonomously in a dynamical and initially unknown environment. Different models have been proposed how the basal ganglia are involved in some cognitive functions (see Joel et al. (2002), Wörgötter & Porr (2005), Vitay et al. (2009) for an overview). O'Reilly & Frank (2006) propose a computational model for working memory (WM). It is based on the architecture of the executive loop linking the dorsolateral prefrontal cortex (dlPFC) with the BG and the thalamus. In this scheme,

the competition between the direct ("Go") and indirect ("NoGo") pathways of the BG is able to selectively open the gate of WM, by enabling bistable units in dlPFC to exhibit sustained activation through the disinhibition of very narrow thalamocortical loops. The executive loop is thus composed of a multitude of parallel channels that are able to store in a hierarchical manner the different events that are needed to perform a task. It is applied to the 1-2-AX psychological task that requires to maintain or forget several kinds of items in WM at different timescales in order to produce the correct response. Despite the great range of functions achieved by the O'Reilly and Frank (2006) model, it has a number of shortcomings. From a computational point of view, a very large timestep (an external event lasts only one timestep) is used. Moreover, the authors use backpropagation learning rules that are biologically implausible. From the functional point of view, the WM is pre-routed and the BG acts like a gate for fixed memory slots.

Part 3: Integrated sensorimotor representation of the 3D reaching space

- ***Joint visuo-motor features in the parietal cortex***

  After solving problems related to lab set-up for studying eye movements and arm movements in depth, partner UNIBO has performed neurophysiological studies in the medial and posterior sectors of parietal cortex, specifically the medial parieto-occipital cortex. In all the previous studies on the medial parietal cortex, the animals performed arm movements and eye movements to targets located always on the same plane, the fronto-parallel one. In the EYESHOTS project, we extended these studies to the 3D space, using fixation points and reaching targets placed in different locations in the monkey 3D space. The data collected and analysed so far show that the sensorimotor transformations regarding arm reaching and vergence eye movements are likely coupled in the posterior parietal cortex of primates. This brain region is able to link perception to action in the 3D space and to provide this information to dorsal stream areas for the purpose of organizing eye and arm actions in the 3D space. An artificial agent endowed with an implicit coupling between its ocular and limb motor systems can more easily take advantage of both proprioceptive and exteroceptive signals in order to interact with its environment and construct an awareness of it.

  All data, both behavioral and neuronal discharges, are already available for the entire consortium. These data are the basis of joint papers already published in Conference Proceedings and Book Chapters, and others submitted or in preparation, to be published in international, peer-reviewed Journals.

  Comparative assessment with the current state of the art: The posterior parietal cortex (PPC) of the primate brain is a crucial node in the organization of actions in peripersonal space. This fundamental role of PPC in interaction with three-dimensional (3D) space has been obvious since the beginning of the '900, when Holmes described spatial deficits after lesions of PPC in humans (Holmes, 1919; Holmes & Horrax, 1919). Since then, many neuropsychological studies reported impairments in the performance of actions in depth following lesions of medial sectors of PPC. Patients suffering from optic-ataxia, for instance, are able to perceive objects in space, but not to interact correctly with them (Perenin & Vighetto, 1988), especially in depth (Brain, 1941).

  Despite these demonstrations of the essential role of human PPC for the correct performance of everyday actions in 3D space, very few neurophysiological work has been done so far to investigate the neuronal coding of action performance in depth. The fact that this matter has been so far neglected is partially due to the intrinsic difficulty of exploring the challenging topic of hand (and eye) actions in depth.

  The study of the neuronal coding of eye movements in depth has been performed so far only in a few studies. In his pioneering studies in the 80's, Sakata (Sakata et al., 1980) explored the influence of eye-position signals on neuronal discharges by varying the position of the fixation target, not only in the frontal plane, but also in depth. Changing fixation point in a frontal plane requires the change in the version of the eyes; changing fixation point in depth requires a change in eye vergence. The data collected by Sakata first demonstrated that neural activity in PPC could be modulated by both

version and vergence angle of the eyes (Sakata et al., 1980). They found that many neurons in area 7a have a specific selectivity for the depth of fixation. Later on, another area of inferior parietal lobule, area LIP, has been shown to carry on signals related to horizontal retinal disparity (hence to depth) and to fixation distance (Gnadt & Mays, 1995; Genovesio & Ferraina, 2004). No functional study has been so far performed to study the coding of eye movements in depth in medial parieto-occipital cortex, that is reported to be sensitive to eye movements in 2D (Kutz et al., 2003) and that is strongly and directly connected with the only 2 areas so far known to be involved eye-movement signals in depth, area 7a and LIP (Gamberini et al., 2009).

- ***Influence of motor and visual components on fragment location***

  Partner WWU has conducted studies on the influence of motor and visual parameters on object localization obtained from saccade adaptation data. The first study revealed a correlation of the visual properties of the stimulus and the saccade generation for the transfer from motor parameters to visual perception. Different types of saccades (reactive, scanning) were tested in connection with different localization stimuli (flashed vs. stationary targets). A further study revealed mislocalization after saccadic adaptation even in a fixation condition. The results have been reported in deliverable D5.3a at month 15 and published in the Journal of Neuroscience (Zimmermann & Lappe, 2009 [J11]) and the Journal of Vision (Zimmermann & Lappe in press [J12]). In the last working period further studies on the spatiotemporal nature of saccadic adaptation were conducted. The time-discrete nature of saccadic motor adjustment was related more closely to the time-continuous perception of space. A strategy to adjust the speed of adaptation to the reliability of the visual errors was discovered. The results are currently under review. The spatial properties of the visuomotor representation of space was determined in an eye position experiment. Motor gain fields were found, which can be compared to the perceptual gain fields already known from different visual areas. The experiment are conducted for inward and outward adaptation and for horizontal and vertical variations in space. The research activity is on track. These results will help in clarifying modeling details for the representation of the reachable space of Task 4.3 (UJI).

  Comparative assessment with the current state of the art: Recent progress in the area of saccadic adaptation focussed on the handling of motor noise for the stable construction of the three-dimensional visual space. Collins et al. (2009) studied the role of motor noise for localization. They found, that if a blank is introduced to the saccadic adaptation paradigm, the subjects are aware of the endpoint error they made. An inclusion of an efference copy in the motor adjustment would drastically improve adaptation. The existence of motor adaptation in the McLaughlin paradigm (McLaughlin, 1967) excludes this possibility. The results therefore give insight in the complex interconnections of visuomotor representations. A different approach to optimize the handling of noise in the visuomotor system is introduced in the model of Wei & Körding (2009). The reproduced various studies by the use of a relevance estimator of visual feedback, which is closely connected to our study on consistency. Munuera et al. (2009) also found an optimized use of efference copy and sensory information in a saccade sequence task.

- ***Integrated visuo-motor representation in reaching tasks***

  Although the use of artificial neural networks in robotics is very diffuse and not at all novel, this has rarely regarded visuomotor transformations involving arm movements, and never the coordinate control of gazing and reaching movements. Moreover, whilst the use of Self-Organizing Maps (SOM) is relatively common (Fuke et al. 2009), the employment of biologically-inspired Radial Basis Function (RBF) networks is virtually unexplored. Indeed, RBF have been applied so far only to the computation of inverse kinematics, alone (Zhang et al. 2005) or together with SOM (Kumar, 2010).

  To the best of our knowledge, compared to the state of the art, only the work of UJI exploits real stereo vision, realizing a coordinated control of vergence and version movements. Moreover, our sensorimotor transformations are bidirectional, so that our system is the first one that learns to gaze towards its hand but also to reach where it is looking at.

Part 4: Cooperative actions in a shared workspace

Partner WWU has collected and analyzed data both in single-subject and in two-subjects settings to explore specific aspects of human behaviour in the combination of allocation of attention and direction of gaze that can be used for prediction in human-robot interaction. The data in single-subject studies were needed for milestone M9.ante at month 18. The milestone was reached as expected. The first single-subject study explored the hypothesis that other's gaze direction can be used to predict the behaviour of the other person. When participants were allowed to see the other's gaze behaviour, they were quicker and more accurate in identifying the locations of the objects that were in the spot of the actor's attention. We showed that other's gaze direction can be advantageously used as a predictive cue. In the second single-subject study, we explored the interconnection between gaze movements and hand movements. By adapting the standard cueing paradigm we measured overt shifts of attention due to different stimuli that contained either gaze and hand information in isolation or in combination. We showed that gaze cueing effects are much stronger when gaze and hand information are presented in conjunction. This result suggests that human do reflexively orient to gaze movements that indicate potentially relevant actions and not merely to every gaze movement. In the third study, we conducted an experiment in which the eye movements were simultaneously measured in pairs of participants involved in a simple cooperative task. The eye movements during the execution of this task exhibited a stereotypical behavior. Prior to the contact between objects both participants directed a saccade towards the partner's object. The timing of these object oriented saccades was found to be modulated by the predictability of the contact location. Therefore, we suggest that the expectations that a human actor has about the cooperation partner can influence the use of attentional resources for the achievement of the final goal.

The results of the experiments are currently prepared for publication and will be reported in deliverable D5.4 at month 27. Subgroup meetings between WWU and UJI took place for discussing issues of integration between human and robot setups with an emphasis on the implementation of an interface that will allow the connection of the eye tracker to the robot.

Comparative assessment with the current state of the art: Nonverbal communication (gestures, face expressions, gaze direction etc.) play a crucial role in both human-human and human-robot communication. Breazeal et al. (2005) showed that the ability of the robot to communicate either implicitly through behavior or explicitly through non-verbal signals improved the effectiveness in human-robot teamwork, where the robot served as a cooperative partner. Yoshikawa et al. (2006) proposed methods of controlling a robot's gaze responsively to its partner's gaze. The results of their experiments confirmed that a robot with responsive gaze provided a human subject with a "feeling of being looked at", which helped to maintain a close face-to-face communication. Mutlu and colleagues investigated how a robot can establish the participant role of its conversational partners using gaze cues (Mutlu et al. 2009a), and explored whether people could make attributions of intentionality in humanlike robots, also mainly through gaze cues (Mutlu et al. 2009b). Since monitoring the gaze of the interlocutor in human communication facilitate partners' mutual understanding, Staudte and Crocker (2008) argue that a robot should also be able to provide the user with non-verbal information concerning the robot's intended meaning or the robot's successful understanding. They found evidence that people look to the objects that the robot refers to linguistically, as well as people look to objects that the robot looks at. We have not found any work in which the robot is treated as a subject of psychophysical/neurological experiments, as in our experimental setup.

----- o -----

From a more integrative perspective, the research activity of the second period addressed more decisively how to incrementally learn or develop articulated and sophisticated ways to interact with the word at our reaching distance. This occurred through:

- The definition and testing (in closed loop and under controlled conditions) of ***vergence-version control strategies with attention effect*** to guide active binocular exploration of the environment (work by K.U.Leuven, UG, WWU)
- The construction of a ***multimodal representation of space through concurrent reach/gaze actions*** (work by UJI, UNIBO, WWU).

Concerning the "vergence-version control with attention effect (VVCA)", the model is composed of 5 parts:

1. Environment simulator, that generate the image stereo pair
2. Robotic head, a kinematic model of the eye movement for a pan-tilt and a tendon-driven binocular head
3. Disparity representation, a model of V1 area for a distributed representation of retinal disparity
4. Object-based saliency, that generate a saliency map to drive the version on an object
5. Eye movements, that generate the controls for the robotic head in order to produce vergence (based on disparity information) and version (based on saliency) movements.

A schematic diagram of the proposed architecture is shown in Fig. 4

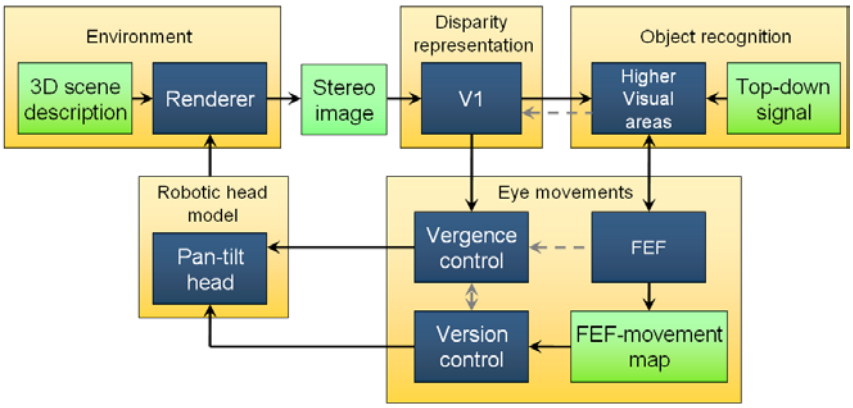


Figure 4. *The block-diagram of the proposed VVCA model.*

The *3D scene description* block contains the information about the peripersonal space observed by the robot (or its model). Depending on the *renderer* this information can be represented in different formats (*i.e.*, MATLAB structure, VRML data). The 3D scenes used are:

- real word scenes, generated from the VRML acquisitions of a 3D laser scanner (work of partner UG)
- synthetic word scenes, generated by a simulator that considers a peripersonal space populated by synthetic textured objects (work of partner K.U.Leuven)

Both can be used for vergence eye movements and for the object-based saliency.
Eventually, the renderer can be replaced by the real stereo setup.

The *Robotic Head Module* (RHM), initially provided by K.U.Leuven, has been replaced by the Simulink head model developed by UG partner. The RHM block takes as input rotational velocities for both eyes and provides the exact position and orientation of the both cameras (eyes) to the renderer. The robotic head is composed of a bio-inspired ocular model and a pan-tilt platforms commonly used in robot vision.
The oculomotor plant is composed of:

- Head block
- Eye block
- EOMs (extra ocular muscles) block

The pan tilt system is composed of:

- Head block
- Pan-tilt block
- Joint velocities block

For each block it has been developed a custom graphical interface (with the MATLAB GUI) to configure the parameters of the block.

The *disparity representation* occurs in the *V1 block* by the response of a population of binocular energy complex cells (work by partner UG), and is used both for the object-based saliency and for vergence eye movements.

The *Higher visual areas* block recognizes and combines orientation and disparity information into more complex structures, which then used in *Frontal Eye Field* (*FEF*) module. After successful recognition the responses are merged over position and saved as the Top-down signal patterns for the up-coming visual search. The object recognition using top-down (attention) signal can search for the specific features independently of their position in another scene. The FEF saliency maps are calculated from higher visual areas and determine the position of the object. The output of FEF consists of a saliency map (*FEF-visual*) and a movement signal map (*FEF-movement*).
The *Eye movements* module consists of two sub-modules that produce the velocity control for version and for vergence movements. *Version control* checks if the information provided by FEF-movement is strong enough to initiate a saccade, and if it is the case, it plans and executes a saccadic motion to the position encoded in the FEF-movement map.
The *vergence control* module has lower priority with respect to version control (*i.e.*, works only when saccade is not needed). Due to the established (and fixed) interfaces with other modules (V1, Robotic Head Model), the vergence control module can easily represents the different vergence strategies developed within the consortium (cooperative work of K.U.Leuven and UG).
First, a double-scale strategy with two different resolutions for the foveal and peripheral portions of the visual field is considered. Then, the algorithm developed for the Cartesian domain will be tested in the log-polar geometry. The MATLAB module for the log-polar mapping developed by UG has been made available to the consortium (http://www.eyeshots.it/private/tools.php).

Concerning the "multimodal representation of space through concurrent reach/gaze actions", the architectural scheme proposed by partner UJI is depicted in Fig. 5. The model integrates visual and oculomotor information available from UG, K.U.Leuven and WWU with proprioceptive/exteroceptive information related to the robot arm (position and tactile). The left side of the schema allows us to transform ocular movements and stereoptic visual information to a body-centered reference frame but also, when

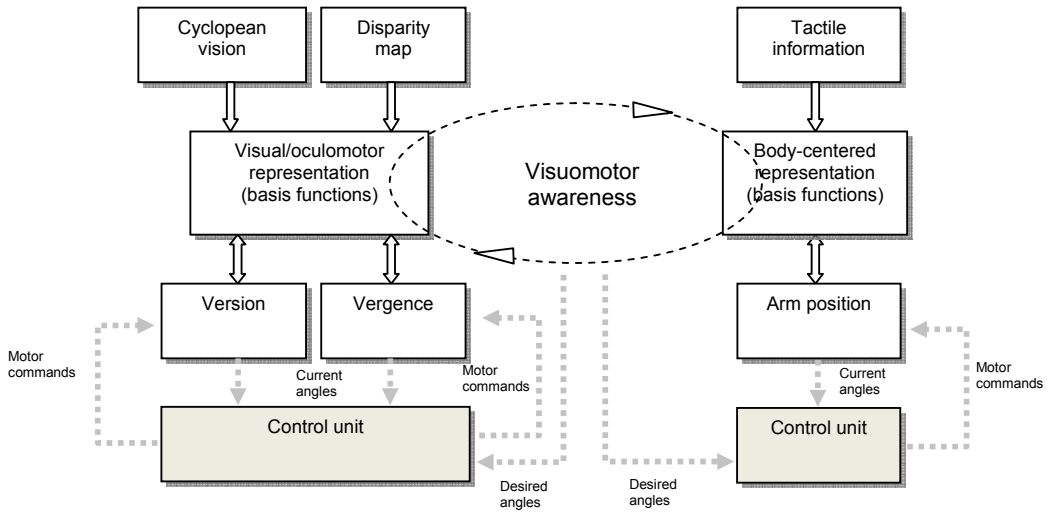


Figure 5. *The block-diagram of the proposed multimodal representation of space through concurrent reach/gaze actions model.*

needed, elicit the eye movements that are necessary to foveate on a given visual target. Log-polar image representations will be used to simulate foveal magnification.
Similarly, the right side of the schema allows us to code the position of objects in the peripersonal space considered as potential targets for reaching movements. The resulting 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.
The model has been designed to embed, as architectural and functional paradigms, the lines of evidences collected by partner UNIBO and WWU on the multimodal tuning properties of the cells in area V6A, and the oculomotor components of visual target localization revealed by saccadic adaptation studies, respectively.

The complex of the two architectures is depicted in Fig. 6 evidencing a hierarchy of distributed representations in which sensorial and motor aspects coexist explicitly or implicitly.

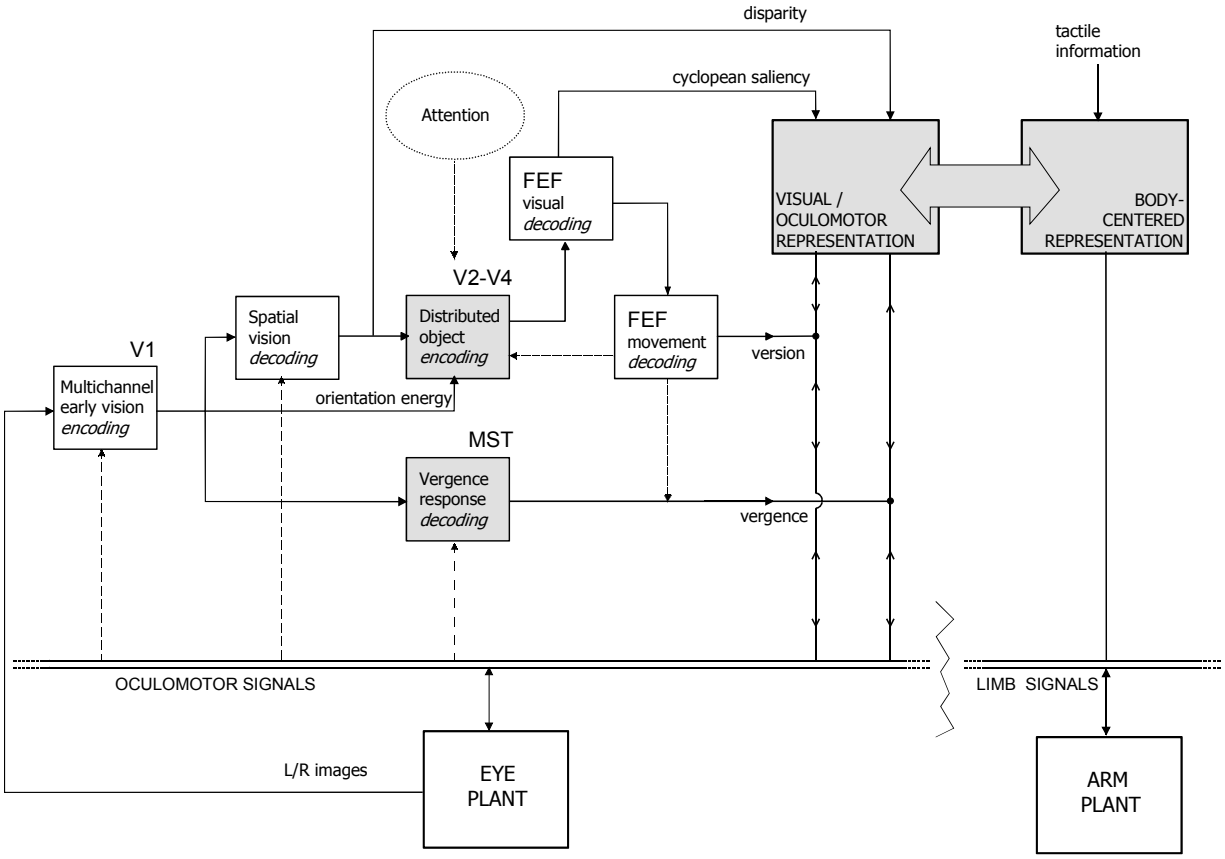


Figure 6. *Schematic diagram of the proposed visuomotor architecture. The processing proceeds through different streams by constructing distributed representation of increasing complexity both for visual interpretation and action. The diagram shows a hierarchy of learning stages (gray boxes) at different levels of abstraction, ranging from the coordination of binocular eye movements (e.g., learning disparity-vergence servos), to the definition of contingent saliency maps (e.g., learning of object detection properties), up to the development of the sensorimotor representation for bidirectional eye-arm coordination. The long-dashed arrows indicate motor influences on the distributed representations.*

On our opinion, this can be considered, **from a methodological point of view**, an interesting intermediate result of the EYESHOTS project. Through such distributed coding, indeed, it is possible to avoid a sequentialization of sensorial and motor processes, that is certainly desirable for the development of cognitive abilities at a pre-interpretative (i.e., sub-symbolic) level, e.g., when a system must learn (binocular)

eye coordination, handling the inaccuracies of the motor system, and calibrate the active measurements of the space around it.
From this perspective, early development of a sensorimotor intelligence is tackled as a problem that jointly involves three concurrent aspects:
Signal processing: by defining the proper descriptive elements of the visual signal and the operators to measure them (cf. the Plenoptic Function (Adelson & Bergen, 1991)).
Geometry of the system and its kinematics: that directly relates to the embodiment concept. The measurement processes on the visual signal is parameterized by specific geometric and kinematic conditions.
Connectionism paradigms: that defines neuromorphic architectural solutions for information processing and representation.

The connectionism paradigm (i.e., hierarchical, distributed computing) is crucial to guarantee accessibility and interaction of the information at different levels of coding and decoding, by postponing decisions as long as possible.

3.2 Workpackage progress

Here we recall the objectives for the tasks of the first period. Quick statements concerning the status are attached; the actual work performed will be detailed for each work package in the WP descriptions below.

WP1: Eye movements for exploration of the 3D space
Task 1.2: Perceptual influences of non-visual cues.
The objective is to analyse the perceptual consequences of specific binocular eye coordination movements and their computational advantages on depth vision and interactive stereopsis.
Scheduling: (month 6-30)
Performed actions: UG developed a theoretical framework to investigate the effects of eyes' torsional components on the perception of the visible surfaces on an observed object (i.e., local patches around the fixation point) for different gaze directions. A preliminary analysis of eye rotations measured in humans and primates (cooperation between UG, WWU and UNIBO).
Results: Under a visuomotor optimization strategy, UG demonstrated that the eyes should move both to maintain the coplanarity of the fixation planes (a property of a tilt-pan system) and to reduce the eccentricity of the rotation. Possible design strategies to embed fixation constraints posed by the oculomotor system into the neural coding and decoding mechanisms of the population of binocular energy units in WP2 have been envisaged.
Status: The work procedes as planned.
Documentation: Deliverable D1.2 (preliminary report)
Publications: Canessa et al., 2010 [J1]
Revised planning: none

Task 1.3: Control of voluntary eye movements in 3D.
The goal of the task is to model the action of the extraocular muscles to achieve correct ocular motions.
Scheduling: (month 6-30)
Performed actions:
- Modelling of the actuation system of the ocular system. Investigation of bio-inspired models of the EOMs and connective tissue surrounding the eye-ball.
- Analysis of actuation techniques for regulating the eye orientation.

Results:
- Parameterization of eye plant for the design and control of a bio-inspired robot eye.
- Definition of a 2D computational model for the definition of the EOM tension to regulate the eye orientation in 3D.

Status:
- Models developed and implemented in simulator (see task Task1.4)

- Complete analysis of mapping EOM static action forces to 3D ocular orientation
- Simulation tests ongoing, and extension to describe transient motions.

Documentation: --
Publications: --
Revised planning: none

Task 1.4: Bioinspired Stereovision Robot System.

This task is focused on the design of a human sized and bioinspired binocular robot system capable to emulate basic ocular movements.

Scheduling: (month 13-36)
Performed actions:

- Development of a dynamic simulator for the comparative analysis of different control strategies and for version and vergence control implemented for difference typologies of robot eye-head systems.
- Study of the design solutions for the implementation of the a bio-inspired binocular head-eye robot. Concept design. Preliminary tests.

Results:

- First release of a dynamic simulator for studing ocular mechanics and visual processing and control techniques (Deliverable D1.4a).
- Concept study of the bio-inspired robot eye.
- Selection of main components for implementation of the eye prototype.

Status:

- First release of the software tested.
- Preliminary tests for integrated image based closed loop control simulation performed.
- Complete integration with VR simulator and control software modules in progress.
- Test rig for actuator and control tests in place.
- Preliminary experiment for assessing actuation accuracy ready to start.
- CMOS camera design in progress
- Complete mechanical design in progress.

Documentation: D1.4a
Publications: --
Revised planning: Following the reviewers' comments (raised at the end of the 1st review meeting, and not included in the written report), on the opportunity of developing a software simulator of the biomechanics of binocular eye movements, the Consortium decided to develop a simulation environment for comparative analysis of bioinspired ocular motions with respect to standard robot eye-head systems (e.g. pan-tilt binocular systems) while maintaining the final objective of implementing the mechatronic anthropomorphic system, An additional effort has been devoted to this task. This extension represents a benefit for the whole project, guaranteeing a first step towards the integration of the different components and the test of the control strategies developed for eyes' movements. The simulator, is now available for integration with the other components.

WP2: Active stereopsis

Task 2.1: Network paradigm for intelligent vergence control

The objective is to develop a convolutional network-based vergence control from a population of disparity-based feature detectors (cooperation between K.U.Leuven and UG). The aim is to learn a vergence motor strategy that, combined with the disparity sparse detectors, optimizes the quality and efficiency of the feature-extraction for the specific tasks.

Scheduling: (month 1-30)
Performed actions: (1) Functional characterization of a cortical-like architecture that implements the dual-mode vergence control; (2) Design of convolutional networks to learn proper disparity-vergence servos directly from examples of the desired vergence behavior.
Status: The work runs as planned. The deliverable D2.1a was submitted on time and the Milestone M5 reached on time.
Documentation: Deliverable D2.1a. Technical meeting notes by Nicolay Chumerin and Frederik Beuth (Chemnitz, 2-9 January 2010).

Publications: Chumerin et al., 2009 [C2], Chumerin et al., 2010 [J2], Gibaldi et al., 2010a [J5], Gibaldi et al., 2010b [C21].
Revised planning: none

Task 2.2: Interactive depth perception.
This task is concerned with the extraction of depth (3D structure) by integrating disparity information across different eye movements. However, when transforming disparity from eye- to head-centric coordinates, the motor part of the EYESHOTS' anthropomorphic head is not accurate enough, therefore, vision is used to improve upon this.
Scheduling: (month 6-36)
Performed actions: (1) Developed a strategy for phase-based disparity estimation in the presence of inaccurate gaze estimates; (2) Robustified phase-difference methods to large orientation disparities; (3) Integrated disparity and epipolar geometry estimation in a coarse-to-fine framework; (4) Extended the method to operate directly in the cortical domain (space-variant log-polar transformation).
Results: Algorithms (available to the consortium as a Matlab software package) for simultaneous disparity and epipolar geometry estimation in the retinal and cortical domain. The algorithms return improved gaze estimates that enable the integration of disparity estimates across eye movements.
Status: The work progresses as planned and even operates in cortical space (log-polar transformation) as requested by the reviewers. The deliverable 2.2a was delivered on time.
Documentation: Deliverable D2.2a, technical note on log-polar mapping based on [J8].
Publications: --
Revised planning: none

WP3: Selecting and binding visual fragments
Task 3.1: Defining visual fragment: object identity
Development of learning algorithms that allow to encode object properties.
Scheduling: (month 1-24)
Performed actions: work by partner WWU. Development of an Hebbian network for learning multiple feature selective units. Expansion to object selective units using stereo vision.
Results: Learning of binocular receptive fields based on Hebbian learning. The network's cells have been compared in detail with physiological observations. Expansion lead to the object selective cells which are tuned to different distances and disparities.
Status: The work has been successfully completed by month 24 (a second journal publication is in preparation). Milestone M7 (Target location for the next eye- movement …) was reached as planned on month 22.
Documentation: Deliverables D3.1a and D3.1b
Publications: --
Revised planning: none

Task 3.2: Selecting visual fragment
Development of dynamic goal-directed attentional selection to bind object properties.
Scheduling: (month 7-30)
Performed actions: Models of receptive field (RF) dynamics before saccade onset. Differences from RF remapping. A model of feature-based attention that has been compared to human experimental data to explore in how far the mechanisms relate to human vision. Selection of object-selective cells via top-down signals.
Results: From the computational point of view, the model of receptive field dynamics that WWU has developed, is consistent with present observations and offers an interesting alternative compared to modelling attention as a simple spotlight. We explored in how far our models capture essentials of human attention.
Status: Continued as planned.
Documentation: --
Publications: Zirnsak et al., in press [J14], Zirnsak & Hamker, in press [J13].
Revised planning: none

Task 3.3: Selecting between behavioral alternatives
Learning of the cognitive control of visual perception.
Performed actions: A first version of the model has been developed and tested.
Results: The model of Basal Ganglia can learn visual target templates according to the task at hand by using only unspecific reward.
Status: Continued as planned.
Documentation: Deliverable D3.3a
Publications: Vitay et al., 2009 [J9], Vitay & Hamker, in revision.[J10]
Revised planning: none

WP4: Sensorimotor integration
Task 4.2: Generating visuo-motor descriptors of reachable objects
The objective is to implement a model of how to generate an integrated sensorimotor representation of objects in the peripersonal space through the practical interaction of an artificial agent with its environment, using visual input and proprioceptive data concerning eye and arm movements.
Scheduling: (month 7-30)
Performed actions:
- Analysis of UNIBO data to orient model formulation and implementation;
- Implementation of visual/oculomotor and oculomotorrm-motor basis function networks which allow bidirectional transformations between retinotopic, head-centered and arm-centered reference frames;
- Adapt the architecture and parameters of the networks to the findings of WP5 regarding V6A and the coding of space; reproduce psychophysiological effects;
- Simulate experiments of learning and define the experimental setup for the real robot;

Results:
- The system is able to accurately learn the transformation between visual, oculomotor and joint spaces, in a way suitable to their application to the robotic setup;
- The system adapts to altered perception and is able to reproduce some effects of saccadic adaptation.

Status: The work proceeds as planned. The model is nearly ready and robotic implementation has begun.
Documentation: Deliverable 4.2a
Publications: Chinellato et al. 2009b [C7], Antonelli et al. 2010 [C20], Chinellato et al. 2010 [J15].
Revised planning: none

Task 4.3: Constructing a global awareness of the peripersonal space
Extend the skills of Task 4.2 to a multiple object setup, in which the agent will simultaneously learn to reach towards different visual targets, achieve binding capabilities through active exploration, and build an egocentric "visuomotor map" of the environment.
Scheduling: (month 19-36)
Performed actions:
- Setup the robotic hardware including arms and anthropomorphic head on a humanoid torso;
- Perform object recognition as done by the ventral stream using also dorsal information for modeling the interaction between pathways;
- Define the final human-robot interaction setup and begin working on the interfaces.

Status: The work proceeds as planned. Multiple object learning framework has been defined.
Documentation: --
Publications: --
Revised planning: none.

WP5: Human behavior and neural correlates of multisensory 3D representation
Task 5.1: Role of visual and oculomotor cues in the perception of 3D space.

The objective of this WP is to collect neurophysiological results to be used to implement computational models developed in other WPs, providing architectural guidelines for the organization of perceptual interactions and the production of artificial intelligent systems able to explore and interact with the 3D world.
Scheduling: (month 1-36)
Performed actions: UNIBO conducted monkey training and 2 electrophysiologial experiments.
Results: One study is experimentally completed, but the analyses are still running; One study is completed and a full paper is in preparation. Results have been proposed to international meetings and shared with the other EYESHOTS partners.
Status: work started and conducted as planned.
Documentation: --
Publications: Fattori & Galletti, 2009 [C3], Breveglieri et al., 2009a [C10], Fattori et al., 2009 [C9], Bosco et al., 2009 [C12], Bosco et al., in preparation.
Revised planning: Although no major deviations from the planned experimental work have been done, some steering of the activities has been decided, in order to follow the suggestions of the reviewers during the first reviewing meeting and to "tune" the experimental activity with the general objectives of the EYESHOTS Project in its executive phase. See page 89-90 for details.

Task 5.2: Link across fragments.
This task is aimed at studying neural correlates of multisensory representation of 3D space obtained through active ocular and arm movements.
Scheduling: (months 1-36)
Performed actions: UNIBO conducted monkey training and 1 electrophysiologial experiment composed of ocular and reaching components.
Results: the study is continuing. Results have been shared with the other EYESHOTS partners and have been used to implement computational models developed in other WPs
Status: work started and conducted as planned. All the Milestones for the period have been reached. M6 (end of monkey training) has been already reached on month 12 (ahead of time).
Documentation: --
Publications: Chinellato et al., 2009a [C8], Chinellato et al., 2009b [C7]
Revised planning: Although no major deviations from the planned experimental work have been done, some steering of the activities has been decided, in order to follow the suggestions of the reviewers during the first reviewing meeting and to "tune" the experimental activity with the general objectives of the EYESHOTS Project in its executive phase. See page 89-90 for details.

Task 5.3: Motor description of fragment location.
The objective of this task is to experimentally determine motor descriptions of eye movements via saccade adaptation to reveal descriptions of fragment locations.
Scheduling: (month 1-36)
Performed actions: WWU conducted five behavioural experiments.
Results: Four studies were completed and are published or in preparation for publications. One study is still ongoing.
Status: Work started and proceeded as planned.
Documentation: --
Publications: Zimmermann & Lappe, 2009 [J11], Zimmermann & Lappe, 2010 [J12], Breveglieri et al., 2009b [C13], Galletti et al., 2010 [J4]
Revised planning: A cooperative effort between UNIBO and WWU on similar experiments on the monkey has been rescheduled for the third period in order to be able to focus in the second period on a cooperative study on attention in V6A between UNIBO and WWU

Task 5.4: Predicting behaviour and cooperation in shared workspace.
The objective of this task is to study specific aspects of human behaviour in the combination of allocation of attention and direction of gaze that can be used for prediction in human-robot interaction.
Scheduling: (month 12-36)

Performed actions: WWU conducted three behavioural experiments, two in single-subject and one in two-subjects settings.
Results: The studies were completed and are currently in preparation for scientific publications.
Status: Work started and proceeds as planned. Milestone M9.ante (Experimental data in single actor setting obtained) was reached as planned on month 18.
Documentation: --
Publications:
Revised planning: none

WP6: Project coordination and management
Scheduling: ongoing
Performed actions: See section 5.1
Revised planning: None

WP7: Knowledge management, dissemination and use, synergies with other projects
Task 7.1: Regular publications of webpages
Scheduling: ongoing
Performed actions: See section 5.2.
Task 7.2: Internal workshop sessions
Scheduling: ongoing
Performed actions: The program of the IURS'09 EYESHOTS Summer School, including both internal and external contributions, allowed an interdisciplinary exchange of knowledge within the consortium.
Task 7.3: External dissemination
Scheduling: ongoing
Performed actions: See section 5.2
Results: We have by now published 23 conference contributions and 21 journal papers.
Task 7.4: Synergies with other projects
Scheduling: ongoing
Performed actions: With respect to the previous reporting period, no other specific actions have been undertaken.
Revised planning: none

WP8: Training, education and mobility
Task 8.1: Literature database
Scheduling: ongoing update of the database
Task 8.2: Student's seminars
Performed actions: Considering he numerous occasion of exchange of knowledge between students, no formal acts have been undertaken.
Task 8.3: Personnel exchange
Scheduling: ongoing
Performed actions: Several visits took place among partners in the reporting period.
Task 8.4: Summer school
Status: completed
Performed actions: The Summer School has been organized (within the scheme of the International UJI Robotics Summer School (IURS) of the UJI partner) and successfully held in Benicàssim (Spain) on September 14-18, 2009.

In the following, we provide a detailed description of the progress of work for each work package -- except project management, which will be reported in section 5.

WP1: Eye movements for exploration of the 3D space

Leader: Giorgio Cannata (UG)
Contributors and planned/actual effort (PMs) per participant:
UG (21/21), UJI (2/1.5) and K.U. Leuven (0/0).
Planned/actual Starting date: Month 1/1

Workpackage objectives

The major goals of the workpackage are the study of ocular mechanics and oculomotor control, for both single eye and conjugate movements, as well as the specification of ocular motion strategies which could improve the capabilities of vision to perceive depth information. In particular, the target is to investigate the role of the ocular mechanics with respect to the strategies implemented by the brain to drive typical biological ocular movements (including saccades and vergence). A second objective is the study of the geometric and kinematic effects of ocular motions on image flow, for supporting the estimation of 3D information from ocular motions. The final goal of WP1 is the development of a bio-inspired stereoscopic robot system capable to emulate the ocular motions to be used during the planned experimental tests.
The starting point of the second year has been related on one hand to investigation of the actuation strategies required to drive the eye motions assuming a tendon driven structure emulating the human eye (Task1.3); on the other hand to address the problem of designing the final robot eye representing the experimental target of this WP.

Progress towards objectives

Task 1.2: ***Perceptual influences of non-visual cues***
The eye with its three DOF could, in principle, assume an infinite number of torsional postures for any gaze direction. Donders discovered that only one torsional position exists for each combination of the azimuth and the elevation angles. Listing's Law, going one step further, states that there is a primary position such that the eye assumes only those orientations that can be reached from this position by a single rotation about an axis in a plane called Listing's plane (Tweed & Vilis, 1990). During convergence, this plane is rotated temporally through an angle Φ_l and Φ_r, respectively for the left and the right eye. These convergent-dependent changes of orientation of Listing's plane have been referred to as the binocular extension of Listing's Law or L2 (Mok et al., 1992). From a functional point of view, Listing's Law can be understood as a limiter of the degrees of freedom of the oculomotor system (when the behavioral situation demands or permits it) that constraint torsional components to reduce rotation (Listing's Law or LL) (Tweed & Vilis, 1990), or to reduce the cyclovergence and restricts the motion of the epipolar lines (Binocular LL or L2) (Mok et al., 1992), thus permitting stero matching to work with smaller search zones. In the spirit of deriving a justification of the Listing's Law with respect to both "motor" and "visual" criteria, our aim was to derive the pair of values for the ϕ angles that allow us to meet some optimality visuo-motor principle, that maximizes vision and motor efficiency (Tweed, 1997). Accordingly, we define a cost function to be minimized that takes into account both the efficiency constraints:

$$F(\Phi_l, \Phi_r) = (1 - \alpha)\, M(\Phi_l, \Phi_r) + \alpha V_1(\Phi_l, \Phi_r) + \beta V_2(\Phi_l, \Phi_r)$$

The first term characterizes the primary position with the role of a "special" position for the oculomotor system. This is described in terms of the eccentricity ε of the rotations of both eyes. A small value for M implies that the eyes do not drift too much away from the primary position. For the visual part of the functional we defined two types of constraints (see Fig.7). With the first term we impose the projection on the horizontal and the vertical meridian retinal line be aligned. This constrain gives rise to specific binocular correspondences in the retinal image planes, which we consider as a "reference situation" invariant with respect to the direction of gaze. The second term imposes the coplanarity of the fixation planes. This second type of constraint was inspired by the works and the ideas of Jampel (Jampel, 2008), who claims that the eyes move without torsion in any direction of gaze.

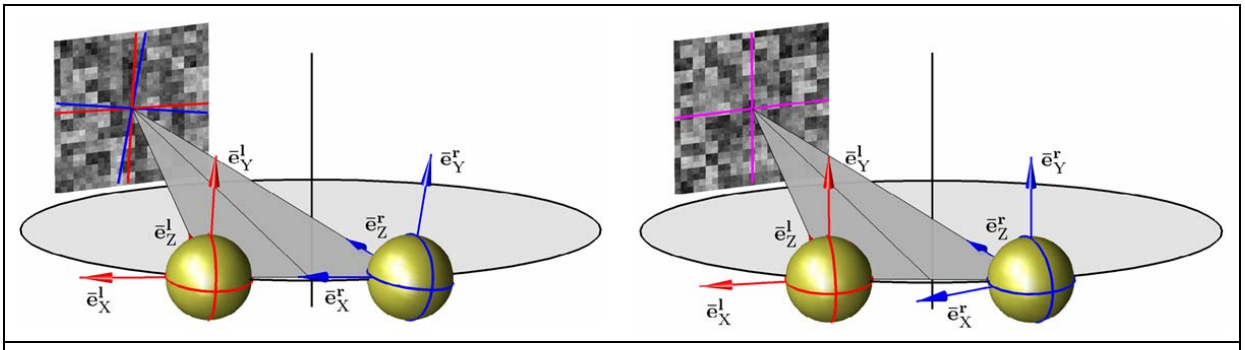


Figure 7. *Three different way to fixate a point in 3D. (Left) Non optimal eyes coordination. The projections are not aligned. (Right) Optimal coordination. The projections are aligned and the fixation planes are coplanar.*

The results, obtained by functional minimization of a weighted combination of the visuo-motor constraints, are not in contrast with the experimental evidences and suggest a continuum of behaviors from far to near vision, also for non-null version conditions (see Fig.8). For any fixed choice of the weights in the cost functional, it is the task that conditions the behavior of the eye movements, moving continuously from a Listing compliant system, during monocular viewing of a distant object, to an Helmholtz- (i.e., tilt-pan) system in binocular near vision. The perceptual advantages for stereopsis have been analyzed in terms of the motion of the epipolar lines. Fig.9 shows that the adopted oculomotor behavior allows us to reduce the search zone within the stereo correspondences have to be found and thus the associated computational cost.

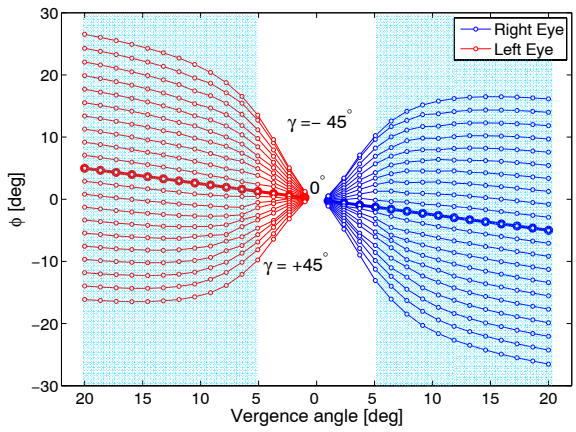


Figure 8. *The rotation angles of Listing's planes as a function of vergence and version. The values of the Φ angles ($Φ_l$ in red and $Φ_r$ in blue) are depicted. It is worth noting that the Φ angles are linear functions of the vergence, but the slope of the curves changes with the version γ. The thick line corresponds to the null-version case (γ = 0°).The near vision regions are highlighted in cyan,, while in white region represents the far vision region.*

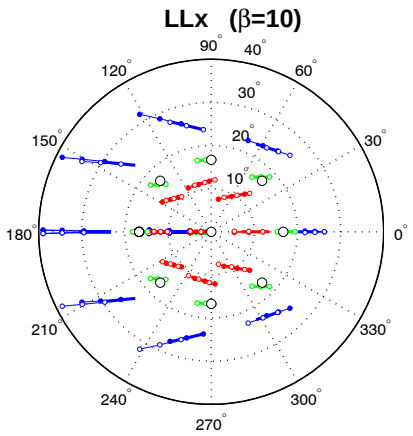


Figure 9. *The resulting epipolar lines for the different strategies obtained from the optimization of visuomotor efficiency. The results are related to a fixed vergence of 20◦ and to different values of version and elevation. The blue, the green and the red families correspond to a version angle equal to −45◦, 0◦ and 45◦, respectively. The filled circle, the no-circle and the open circle lines correspond to an elevation angle equal to −45◦, 0◦ and 45◦, respectively. The reference retinal points are indicated by large open circles.The situation is characterized by functional parameters α = 0.95, and β = 10.*

Following a similar approach to (Hansard & Horaud, 2010), we systematically quantified the magnitude of Helmholtz torsions and the degree of failing in terms of motor efficiency for the different systems that result from the functional minimization. The used values of vergence correspond to the typical range of the human peripersonal space (30 – 85cm, for an interocular distance of 6cm) for close object inspection and manipulation. The results obtained for torsions are shown in Table 1. The last row refers to the torsions predicted by the Listing's Law when the system fixates at infinity ($\upsilon = 0$), and can be directly compared to the ones derived in (Hansard & Horaud, 2010) and reported below the table in italics. As the vergence increases, the measured averaged torsions predicted by our generalized visuo-motor constraint diminish up to about 30%. It is worth noting that, on the contrary, for LL and L2 the amount of torsion remains basically unaffected by vergence. The results obtained for the eccentricity are shown in Table 2, where a column relative to the Mean Absolute Percentage Error (MAPE) is added. The variation of eccentricity is extremely small for all the situations considered. The values in italics correspond to the average values of $\left|\Delta\varepsilon_{\alpha}^{\beta}(\upsilon)\right|_{0}^{\varepsilon}$, over the different vergences, for LLx. For a direct comparison, the average values obtained for L2 and a tilt-pan system are also reported. Considering the order of magnitude of the values, we can conclude that the motor efficiency, at least in terms of the minimal rotation eccentricity, is a negligible issue to explain the binocular eye movements predicted by L2.

LLx ($\beta = 10$)

$\left\|\tau_{.95}^{10}(\upsilon)\right\|_{0}^{\varepsilon}$	$\varepsilon = 15°$	$\varepsilon = 30°$	$\varepsilon = 45°$	$\varepsilon = 60°$	$\varepsilon = 75°$
$\upsilon = 11°$	0.0698°	0.2858°	0.7813°	1.7941°	3.6709°
$\upsilon = 6°$	0.1408°	0.6113°	1.7136°	3.5863°	6.4975°
$\upsilon = 4°$	0.2530°	1.0057°	2.4480°	4.5405°	7.8998°
$\upsilon = 0°$	0.3390°	1.2951°	2.9390°	5.4035°	8.9134°
	0.314°	*1.280°*	*2.978°*	*5.596°*	*9.630°*

cf. (Hansard & Horaud, 2010)

Table 1. *Absolute values of the torsion magnitude $\tau_{\alpha}^{\beta}(\theta,\psi)$ averaged over increasingly large regions of the viewing sphere and for different values of the distance vergence υ. The optimization parameters used are $\alpha = 0.95$, $\beta = 10$. The row in italics below the table shows the values obtained by Hansard and Horaud (Hansard & Horaud, 2010), which can be directly compared to those we obtained in the null-vergence case ($\upsilon = 0$).*

$\left\|\Delta\varepsilon_{.95}^{10}(\upsilon)\right\|_{0}^{\varepsilon}$	$\varepsilon = 15°$	$\varepsilon = 30°$	$\varepsilon = 45°$	$\varepsilon = 60°$	$\varepsilon = 75°$	**MAPE**
$\upsilon = 11°$	0.0087°	0.0426°	0.1178°	0.2349°	0.3836°	0.4684%
$\upsilon = 6°$	0.0040°	0.0188°	0.0392°	0.0631°	0.0888°	0.1184%
$\upsilon = 4°$	0.0012°	0.0046°	0.0082°	0.0175°	0.0190°	0.0267%
	0.0046°	*0.0220°*	*0.0550°*	*0.1051°*	*0.1638°*	**LLx**
	0.0029°	*0.0058°*	*0.0087°*	*0.0117°*	*0.0149°*	**L2**
	0.0092°	*0.0610°*	*0.2046°*	*0.5123°*	*1.1100°*	**Tilt - Pan**

Table 2. *Absolute values of the eccentricity error $\Delta\varepsilon_{\alpha}^{\beta}(\theta,\psi)$ averaged over increasingly large regions of the viewing sphere and for different values of the distance vergence υ. The optimization parameters used are $\alpha = 0.95$, $\beta = 10$. In the last column the Mean Absolute Percentage Error is reported. The three rows in italics below the table show the corresponding averages (over the different vergences) for the LLx, together with those obtained for L2 and a tilt-pan system.*

Task 1.3: ***Control of voluntary eye movements in 3D***

Listing's Law represents an important geometric (and kinematic) constraint to describe a large class of ocular motions. Under reasonable assumptions it is possible to characterize the ocular mechanics as a function of a single major parameter corresponding to the angle β formed by the position vector of the insertion points of

the extra-ocular muscles (EOMs) on the eye-ball with respect to the gaze direction (assuming the eye at rest in its *primary position*). The EOMs are modelled to be routed through pulleys located behind the globe. If the angle formed by the position of the pulleys with respect to the opposite of the gaze direction (assuming the eye at rest in its *primary position*) is also β then it can be formally proved that for any possible force generated by the EOMs the motion of the eye will respect Listing's Law. This results implies that, in principle Listing's Law could be enforced without any explicit action implemented at brain level.

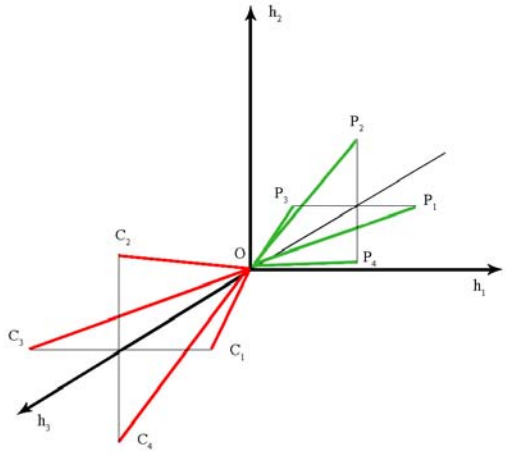


Figure 10. *See text.*

Furthermore, due to the simplicity of the eye-muscles system configurations, the β term represent the main design parameter required to address the implementation of a bio-inspired robot eye.
Figure 10 shows the ideal relative position of the insertion points c_i and the pulleys p_i in the primary position. For the constraint posed by angle β the configuration is fully symmetric.
It is worth noting that even though β is a design parameter its value affects the actual ocular workrange of the eye. It is possible to show that for $\beta = 60^o$ the ocular range is maximized (without causing singularities in the actuation system according to the proposed model), and it is interesting to observe that this value of β is very close to the corresponding average angle observed in humans where $\beta \cong 55^o$. For this reason the latter is the value fixed for the final robot implementation.
From the previous discussion it emerges that accord to the *β-model* any force generated by the EOM generate Listing compatible motions, however it is not specified which forces could drive the eye to fixate in given 3D direction.
To address this problem it is necessary to recall that Listing's Law (in its differential forms also known as *Half Angle Rule* and *Generalized Half Angle Rule*) states that both angular velocity and acceleration of the eye lay on a rotating plane whose orientation forms an angle which is (at any time) half of the actual rotation angle θ of the eye. Let us denote the normal to the rotating plane as n_ω. It is possible to show that the net torque generated by the EOM during Listing compatible motions can be expressed by the following formula

$$\tau = n_\omega \times \sum_{i=1}^{4} \gamma_i \, r_i$$

where

$$\gamma_i = R \, f_i \, \frac{1}{|n_\omega \times r_i|}$$

where R is the radius of the eye-ball, $f_i > 0$ is the actual force generated by the i-th extra-ocular muscle, and r_i are the positions of the insertion points in the rotated eye position.

It is worth noting that τ is in practice orthogonal to both n_ω and a convex linear combination of r_i. The geometry of the system in the rotated position of the eye is shown in Fig. 11.
Recalling that under Listing's Law constraint the finite eye rotation axis v belongs to the h_1-h_2 plane (see Fig. 11) then the configuration shown can be transformed into an head-fixed one by performing a rigid rotation about v of $-\theta$ radians the above formulas can be re-written as:

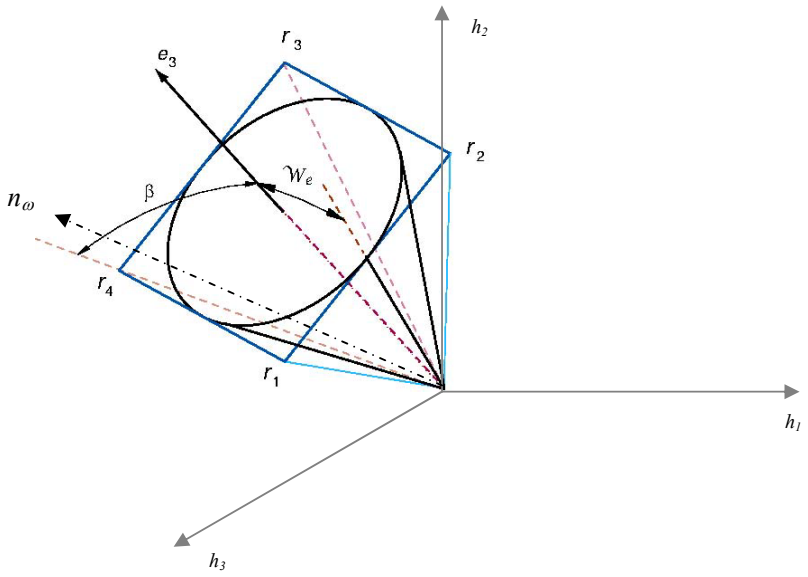


Figure 11. *See text.*

$$\tau = n_\omega^R \times \sum_{i=1}^{4} \gamma_i \, c_i$$

where

$$\gamma_i = R \, f_i \, \frac{1}{\left| n_\omega^R \times c_i \right|}$$

The two formulas above involve this time head fixed terms except for the vector $n^R{}_\omega$ which is rotated (with respect to a head fixed frame) by $-\theta/2$ radians. Furthermore, simple trigonometry allows us to express all the vectors in the above formulas in terms of quantities described in the h_1-h_2 plane.
Assume now to regulate the rotation of the eye to a desired fixation direction, then the coefficients $\gamma_i > 0$ must be chosen. It is worth noting that at steady state the following static equations must hold:

$$\tau - k\theta \, v = 0$$

where $-k\theta \, v$ is the elastic restoring torque due to the connective tissue surrounding the eye-ball. Therefore, the coefficients $\gamma_i > 0$ must satisfy the following equality.

$$k\theta \, v = n_\omega^R \times \sum_{i=1}^{4} \gamma_i \, c_i$$

The computation of the coefficients can be performed using different numerical techniques which are currently under comparison in order to better exploit the most relevant links related with visual data coming

from the vision modules. It is worth noting that due to the symmetry of the placement of the insertion points c_i then the actual structure of the coefficients can be decoupled as follows:

$$\gamma_i = \gamma_i^p + \gamma_i^o$$

where the first coefficients satisfy the above static equations, while the second coefficients allow to satisfy the following equality

$$n_\omega^R \times \sum_{i=1}^{4} \gamma_i^o \, c_i = 0$$

Therefore the coefficients γ_i^o allow to modulate the tension of the EOM without affecting the target eye orientation.
Simulation tests have shown the effectiveness of the proposed model.

Task 1.4: ***Bioinspired Stereovision Robot System***

The activity related to the electro-mechanic design of a bio-ispired robot eye, has been extended with the development of a dynamic simulator for the comparative analysis of different control strategies and for version and vergence control implemented for difference typologies of robot eye-head systems. In particular an accurate dynamic model of a bio-inspired robot eye corresponding to the specifications of the actual robot prototype, target of this Task, has been developed. For comparison different dynamic models of pan-tilt binocular systems have been implemented in order to model ideal behaviour (i.e. intersecting axes) as well as more realistic operational conditions by modelling typical geometric uncertainties (as axes offset and offset rotations). The simulator has been designed as a stand alone Simulink™ toolbox and supported by several graphic user interfaces (GUIs) in order to simplify the configuration of the simulation environment. As a matter of fact the configuration of the single bio-ispired eye involve several parameters related to the geometric, dynamic characteristics of the globe, to the geometric and dynamic characteristics of the extra-ocular muscles (EOMs) and to the characteristics of the connective tissue surrounding the globe. A large set of parameters have been also used to characterize the general model of the pan-tilt binocular systems. The support of GUIs should simplify the set-up of a simulation environment also to a non expert user: default parameters related to literature data have been included to this aim.
The simulator has been also designed to be integrated with the Virtual Reality image synthesizer developed during the First Period of this project by UG.
Preliminary functional tests have been performed in March 2010 aiming at the integration of the software for vergence and version control developed by K.U. Leuven and WWU.
The second activity performed during this period has focused on the study of the design solutions for the implementation of the a bio-inspired binocular head-eye robot. The main specifications of the system are:

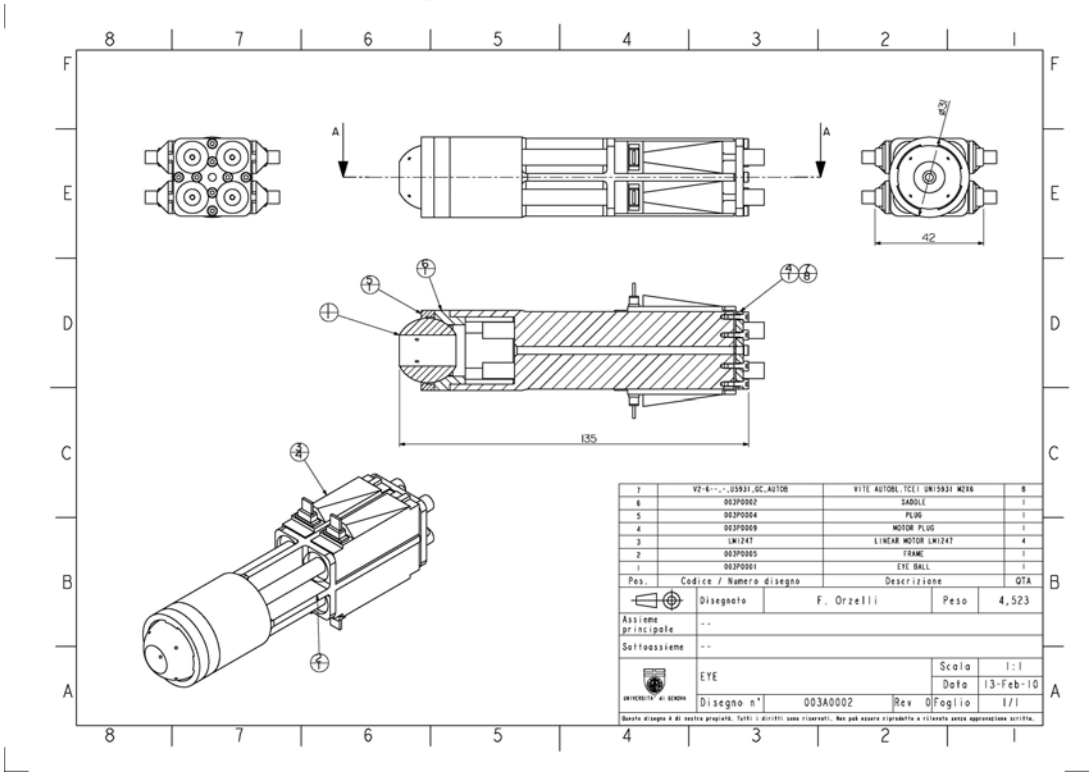
- Human like eye ball geometry: the design aims at the development of an eye ball with a diameter in the range 25-30mm which is quite close to the human size. The only major limitation is the possibility to integrate into the eye-ball a good quality CMOS camera for visual feedback, with USB2 interface. As no commercial devices have been found, a custom design device is currently being designed. As a backup solution if development time should become too long is to adopt an analogue camera with RGB or composite video output. The eye-ball will be made of low friction co-polymer and precision machined in order to host the miniature CCD camera.
- Tendon driven actuation: four recti muscles emulated by cables driven by four linear brushless actuators. The selected motors are direct drive linear actuators (Faulhaber LM 1247-040-02). These devices are relatively small to allow a reasonably compact design. Furthermore, they are direct drive and should allow us to implement direct force control through current feedback in order to emulate the EOM dynamics. Tests to validate this design solution are in progress. Alternative design solutions should involve the adoption of linear position sensors or direct coupling of the moving shaft of the motor to a load cell. Both these solution would increase the size of eye and significantly its final cost. Tendons will be made of nylon or nylon coated stranded miniature steel cable (CarlStahl Technocables). Commercial servo drives (ELMO Whistle EL1313 have been selected to limit the risk of late system development.

- Capability of implementing Listing's Law: this will be achieved by appropriate routing of the tendons and placement of the insertion points on the eyeball. The models for the ocular motions, also used in the simulation environment specify a precise placement of the insertion points on the eye ball with respect to the pulleys constraining the motion of the EOMs around the globe. The basic geometric model assumes an ideal pointwise turning point for the tendons. A design solution to relax this requirement has be found based on the precise machining of the eye-ball support.

The following table summarizes the major expected features of the robot eye under development.

	Target Design	Notes
Eyeball diameter	25-30 mm	
Ocular workrange	±45 deg	
Eyeball speed	> 900 deg/sec	Estimated (w/o control and actuators cross coupling, SF = 8, t.b.a.)
Eyeball acceleration	> 7000 deg/sec^2	Estimated (t.b.a.)
Eyeball materials	Low friction elastomers	Very low inertia (w/o embedded camera)
Sensors	Current feedback Linear encoders	Under test Still to be evaluated
Tendons	Nylon+Steel	Under test
Eye control system	Not Embedded	Commercial servo drives

The following figure sketches the concept design of the robot eye with the currently selected actuators.



Deviations from the project workprogramme

Development of simulation environment for comparative analysis of bioinspired ocular motions with respect to standard robot eye-head systems (e.g. pan-tilt binocular systems). After the first project review, the Consortium has agreed to develop a simulation tool for the assessment of the overall performances of the control algorithms and vision processing techniques. This tool, though not part of the original workplan, represents a benefit for the whole project and guaranteeing a first step toward the integration of the different components making possible the seamless analysis of both vision and control algorithms.

WP2: Active stereopsis

Leader: Marc Van Hulle (K.U.Leuven)
Contributors and planned/actual effort (PMs) per participant:
UG (9/10) and K.U.Leuven (12/19)
Planned/actual Starting date: Month 1/1

Workpackage objectives

This Workpackage is devoted to the specialization of disparity detectors at different levels in a hierarchical network architecture to see the effect of learning (higher-order disparity detectors) in the extraction of the binocular features stereo and stereomotion. A vergence motor strategy is learned that, combined with the sparse detectors, optimizes the quality and efficiency of the feature-extraction for the specific tasks (guided by the attention signal). The second task is concerned with the extraction of depth (3D structure) by integrating disparity information across different eye movements. However, transforming disparity from eye- to head-centric coordinates, but also estimating disparity (and controlling vergence), relies on accurate calibration information (in terms of the relative orientation of the eyes), which is not feasible with Eyeshots' anthropomorphic head, due to the limited accuracy of its motor system, therefore, vision is used to improve upon this.

Progress towards objectives

Specific progress on the tasks worked is reported as follows.

Task 2.1: ***Network paradigm for intelligent vergence control.***
As a starting point, at the end of first year we have demonstrated that, by a proper read-out strategy of the disparity energy population responses, one can obtain a specialization of disparity detectors for the control of vergence. Specifically, the read-out weights were determined to obtain a desired set of disparity-vergence responses on which to base a 'dual-mode' vergence control mechanism: a ''LONG'' fast signal is enabled in the presence of large disparities, whereas a ''SHORT'' slower signal is enabled in the presence of small disparities. In the second year, the research activity (in strong cooperation between K.U.Leuven and UG) focused on (1) the functional characterization of a cortical-like architecture that implements the dual-mode vergence control, and (2) the design of convolutional (linear/non-linear) networks to learn proper disparity-vergence servos directly from examples of the desired vergence behavior
Specific attention has been devoted to analyze and to cope with the effect of the vertical disparity when the gaze deviates from the straight-ahead position.
All the proposed approaches developed in Task 2.1 were designed to provide as an output the rotational eye velocities for their direct use as input for the ROBOT-EYE simulator developed in WP1.

Step 1: Impact of vertical disparity on the dual-mode vergence control
We extended the dual-mode cortical model for the control of vergence, based on a population of V1-like disparity detectors (Gibaldi et al, 2010 [J5]), to work not only with the gaze straight ahead, but also at different directions, tackling the effects of vertical disparity. The extended model is able to provide vergence control which is almost unaffected by the vertical disparity, thus able to produce a precise and stable fixation at any binocular azimuth and elevation.

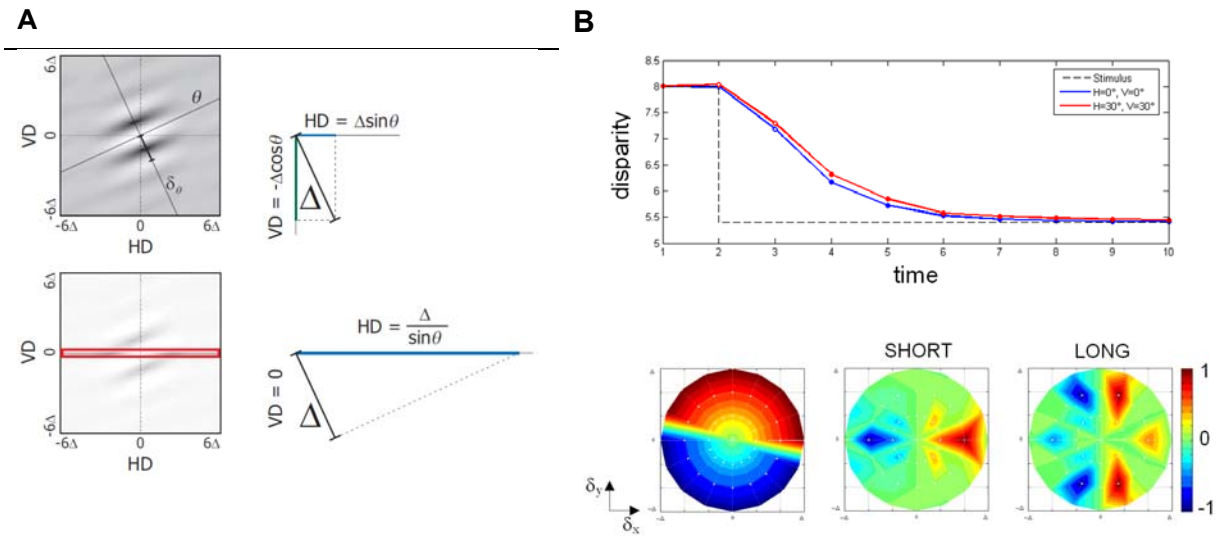


Figure 12: (A) *Each complex cell is, by construction, tuned to an oriented disparity δ_θ i.e., each cell is jointly tuned to horizontal (HD) and vertical (VD) disparities. (Top): For each oriented disparity, its contribution to the HD and VD is calculated by projections on the horizontal and vertical lines. (Bottom): By assuming VD= 0, the orientation of the RF is used as a degree of freedom to extend the sensitivity range of the cell to horizontal disparity stimuli (HD).* ***(B)*** *(Top) Vergence response to a step variation of disparity for a straight-ahead gaze (blue line) and for a tertiary position (red line). (Bottom) Distribution of the weights for the read-out of different information from the population of oriented disparity detectors: disparity estimation (left), and SHORT and LONG control signals.*

Supposing the eyes move following the Helmholtz coordinate system like a standard tilt-pan stereo head, the vergence control at a fixed gaze divides symmetrically on both eyes, regardless the elevation angle. Thus, thanks to the geometry of the system, the models are able to provide an effective vergence control at any binocular azimuth and elevation.

From a modelling point of view, we further investigated the effect of stiumulus vertical disparity, considering that the population of V1-like disparity detectors is jointly tuned to both horizontal (HD) and vertical (VD) disparity. The extended controllability range that characterizes the LONG response in the proposed dual-mode approach is indeed attained by mapping the 2D vector disparity feature space of the oriented binocular energy cells into a 1D space of the projected horizontal disparities, where the orientation θ plays the role of a parameter.
More precisely, by assuming VD = 0, the dimensionality of the problem of disparity estimation reduces to one, and the orientation of the receptive field is used as a degree of freedom to extend the sensitivity range of the cells' population to horizontal disparity stimuli (see Fig. 12). Analyzing how VD tuning (or, from a different perspective, the stimulus VD) does effects the short-latency disparity-evoked vergence and, consequently, the performance of the servos is therefore a critical issue. Indeed, with the straight-ahead gaze, the average VD around the fovea is close to zero, but at different directions, it arises to sensitive values. Through extensive simulations we verified that, although the control yields stable fixations at any binocular azimuth and elevation, it still exhibits a bias on VD. The reason behind this is that the major contribution for horizontal vergence derives from cells mostly tuned to VD, while their responses are weighted to strongly reduce the dependence on VD within its variability across the tested gaze directions. Analyzing which kind of cell influences the horizontal vergence response, it turns out that the most effective are not those mainly tuned to HD, but there is a predominance of those tuned more to VD, and their responses are weighted in such a way that the dependence on VD is strongly reduced in a certain range. As reported in humans (Yang et al., 2003), the model yields the strongest responses when the VD is close to zero, whereas the responses decline as the VD increases. A contribution on the results of this analysis is currently submitted to the European Conference on Visual Perception 2010 [C21].

Step 2: Learning eye vergence control from a distributed disparity representation

Modular framework for vergence control

We consider a stereo setup consisting of a model of fixed robotic head with a pair of eyes (see Fig. 13). The task is *to estimate and then to maintain the vergence angle which brings the fixation point (along the gaze direction) onto surface of the observed object*. Additionally, the model should work without explicit computation of the disparity map, but extract the vergence control signal from the response of a population of disparity tuned complex cells, the actual gaze direction **g** and the actual vergence angle α.

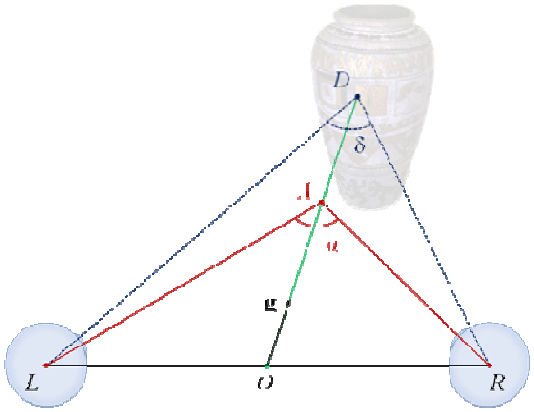


Figure 13: *The geometry of the robotic head model. L and R are the centers of the eyes, O is the middle-point of the baseline LR; A is the actual fixation point and α is actual vergence angle; D is the desired fixation point and δ is desired vergence angle; (unity) vector* ***g*** *depicts the gaze direction.*

With the use of a modular framework (see Fig. 14), we have investigated two neural models for vergence angle control.

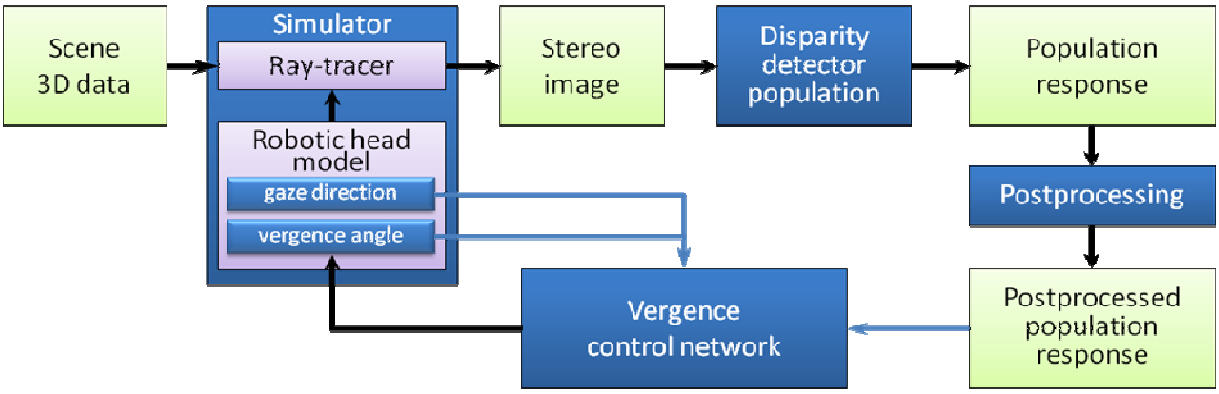


Figure 14: *The scheme of the framework used for VC model training and testing.*

The first model, referred to as the *linear vergence control network* (linear VC-net), obtains a disparity-vergence response by a linear combination of the pooled population response (see Fig. 15). In the second model (see Fig. 16), the vergence angle is controlled by the *convolutional neural network* (convolutional VC-net).

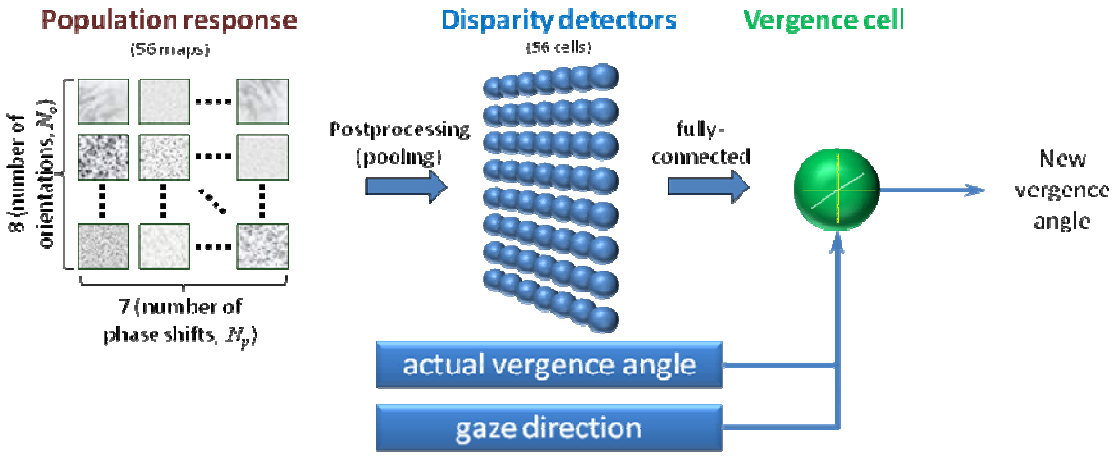


Figure 15: *Linear vergence control network and its inputs.*

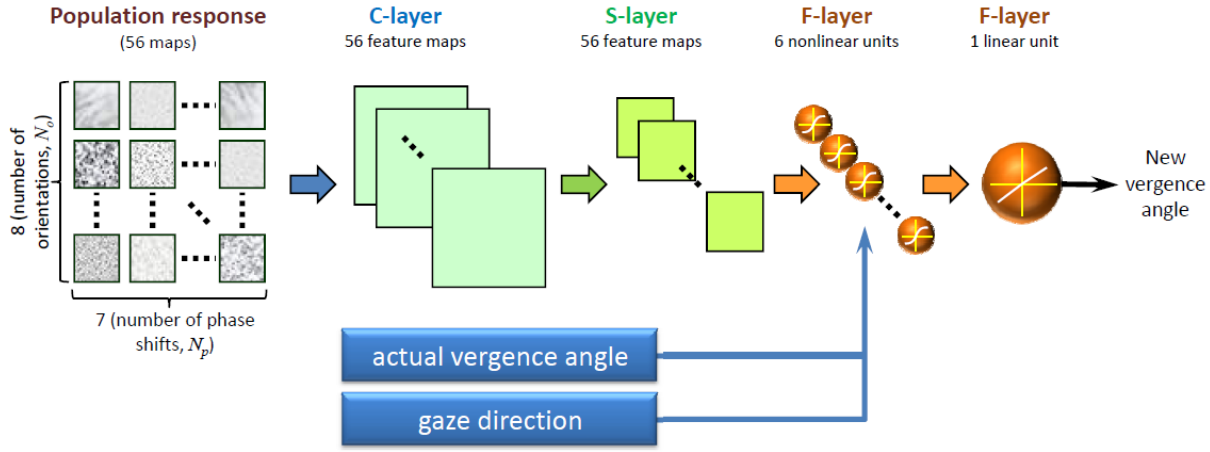
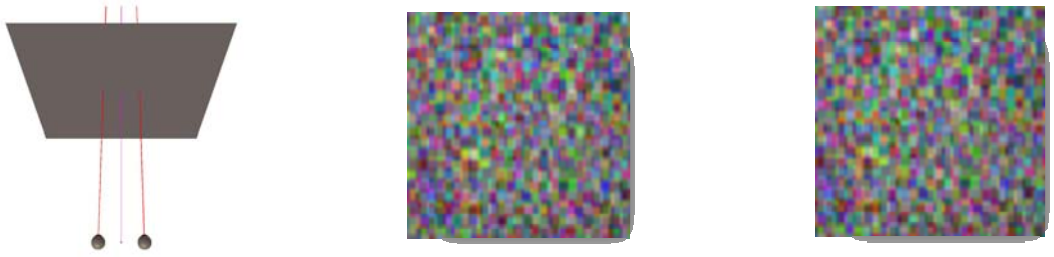


Figure 16: *Convolutional vergence control network and its inputs.*

Both models work in a closed-loop manner, and do not rely on any explicitly computed disparity, but extract the desired vergence angle from the postprocessed response of a population of disparity tuned complex cells (Chessa et al, 2009a [C1]) the actual gaze direction, and the actual vergence angle. Both networks have been tested in two different scenarios: simplified scenario (the gaze direction of the robotic head is always orthogonal to its baseline and the stimulus is a frontoparallel plane, thus, also orthogonal to the gaze direction, see Fig. 17) and the general case scenario (all restrictions on the orientation of the gaze, as well as the stimulus position, type and orientation, are dropped, see Fig. 18).



(a) (b) (c)

Figure 17: *Simplified case scene example (a) and rendered left (b) and right (c) views. The images have the same (low) resolution exactly as it was used in simulations.*

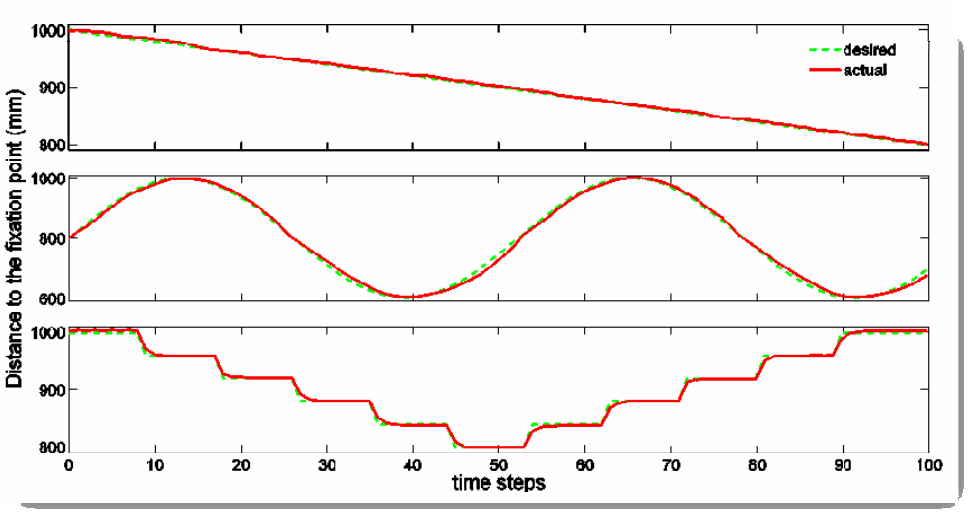
(a) (b) (c)

Figure 18: *General case scene example (a) and rendered left (b) and right (c) views. The images have the same (low) resolution exactly as it was used in the simulations.*

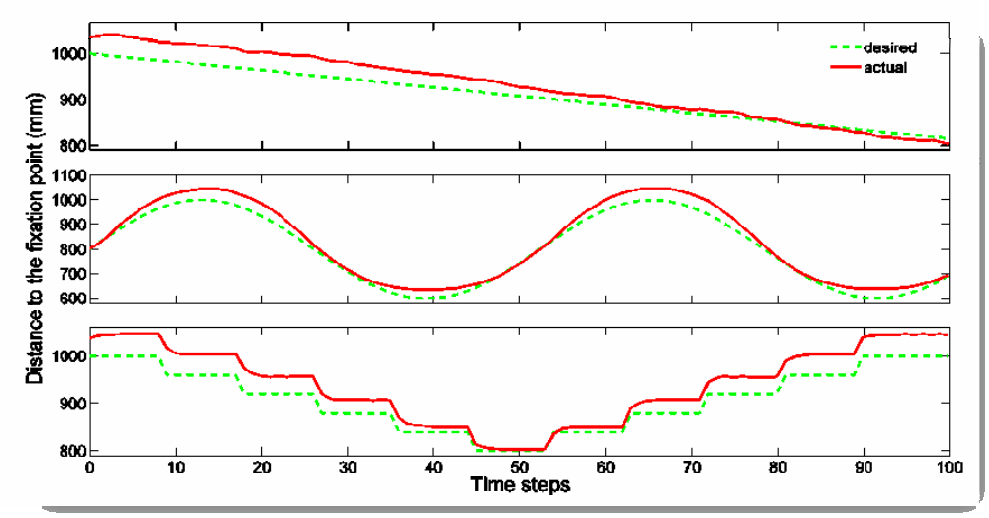
Results and Conclusions

Three standard tests (ramp, sinusoid and staircase) were carried out for the simplified as well as the general case. The typical results of the performance, measured in terms of distance to the fixation point, are shown in Fig. 19. Comparing the performance of the two networks led us to conclude that the convolutional network, although it has slightly larger inertia, is able to handle more accurately the general case tasks than the linear one. On the other hand, vergence control based on the convolutional network is much more computationally expensive than the linear based. Although the model only resorts to a population of neurons in a single scale, we demonstrated that, using a convolutional network, accurate and fast vergence control can be achieved in a closed loop, for different orientations of the gaze.

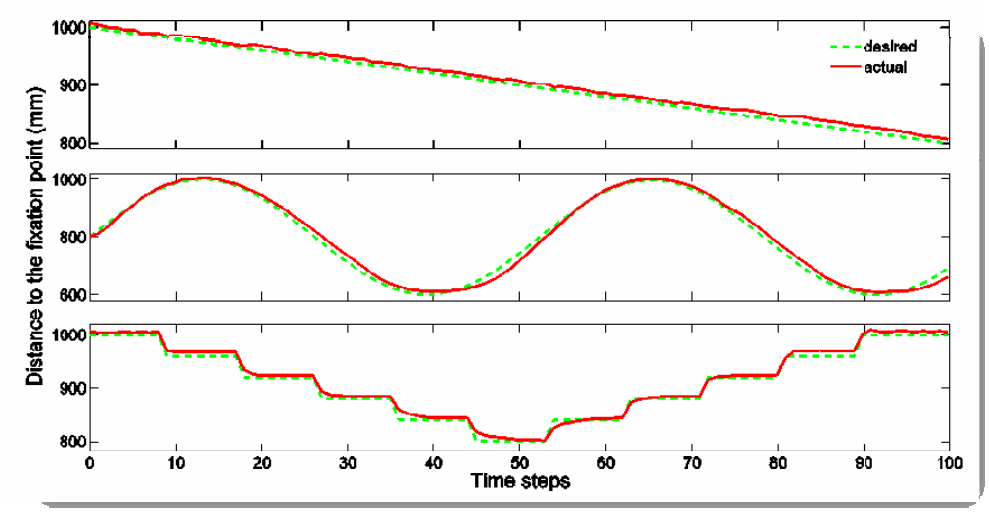
This work has been presented at the International Conf. ISABEL'09 and it is described in a journal article, jointly prepared by K.U.Leuven and UG. (Chumerin et al., under review [J2]).



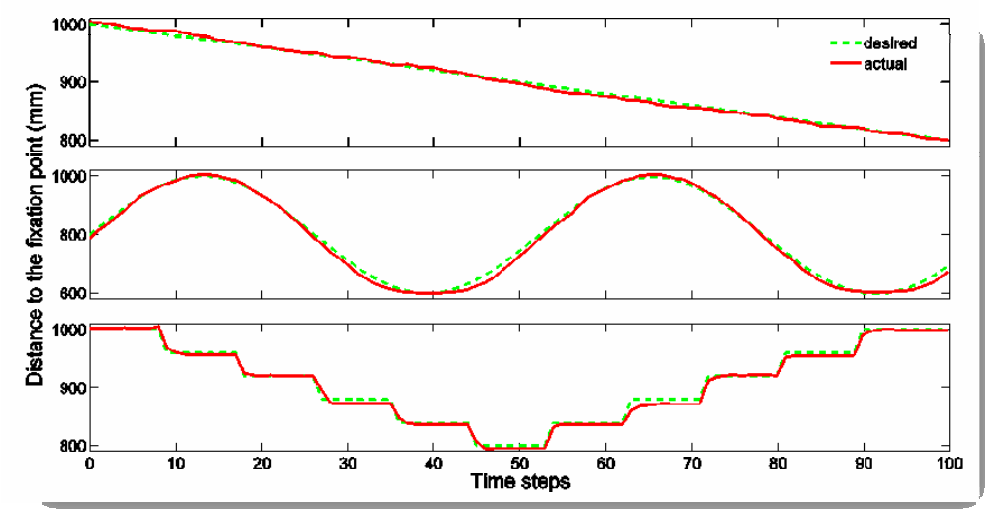
(a) Linear VC-net, simplified case scenario.



(b) Linear VC-net, general case scenario.



(c) Convolutional VC-net, simplified case scenario.



(d) Convolutional VC-net, general case scenario.

Figure 19: *Typical examples of the depth-based performance plots for a linear VC-net in the simplified (a)/general case (b) scenarios and convolutional VC-net in the simplified (c)/general case (d) scenarios.*

Future Extensions

The disparity-vergence responses of the output vergence cell of the convolutional network will be derived for two different gaze directions and systematically compared with their counterparts for the dual-mode model. The first gaze direction is straight ahead in order to have no influence from VD, whereas the second one is at a non-null elevation and azimuth angle in order to verify the influence of VD.

Moreover further generalization of the network paradigm will be explored, also with the aim of including (i) dynamic (i.e., spatiotemporal) disparity tuning, and (ii) attentional signals (based on object properties) that might guide intentional exploration of the selected object (preliminary steps towards this last issue have been moved by defining a proper simulation framework to conduct first closed-loop experiments in controlled situation).

Task 2.2: ***Interactive depth perception.***

On the basis of the strategy for robust interactive depth perception outlined in the previous period, two algorithms have been developed, and their implementations made available to the consortium as a software module (see Deliverable 2.2a). The first algorithm operates directly on images received from the cameras (the retinal domain). It exploits initial gaze (calibration) estimates to improve disparity estimation using a variety of warping mechanisms operating in the spatial and orientation domain. These improved disparity estimates are then, in turn, used to improve the calibration information. The second algorithm uses the same concepts but operates in the cortical domain, on images transformed using a space-variant log-polar mapping. The performance of these algorithms has been evaluated in different configurations (retinal, cortical, close to vergence, and far-away from vergence), and compared to standard algorithms.

The proposed methods can operate together with the vergence mechanisms presented in Task 2.1 in various ways. Improved calibration estimates can feed directly in the convolutional network for vergence control, but can also modulate the weights of the mechanism that integrates the population responses into the vergence control. The methods can also be configured to operate in the periphery only, which is useful for steering

vergence to locations memorized in terms of motor information (WP1). In this situation, there is no need for small foveal disparities. The corrected gaze information is returned in the form of a corrected fundamental matrix. This is used for the coordinate transforms in Task 2.2, but also for the fine motor control in WP1 that is required for the interactive exploration of the fragment. This kind of feedback from the visual system is important for the actual control of the robot eyes in Task 1.4, the requirements of which have been discussed in Deliverable D1.1.

In the remainder we give a brief overview of both methods and some results. For more details and full algorithmic descriptions, see Deliverable 2.2a.

Auto-calibration in the Retinal Domain

Both the retinal and cortical domain methods use phase differences for the estimation of stereo correspondence in the absence of precise calibration information. To estimate possibly large 2D correspondences, the responses from a multiscale and multi-orientation Gabor filterbank are used. Although this requires a computationally intense (but data-parallel) filtering step, the matching itself is effortless (*cf.* gradient-based methods in optical flow). To avoid having to re-filter the images while gradually improving the calibration, we do not actually rectify the images, but rather adjust the read-out of the filter responses. For this purpose, we have developed a method to compensate for large orientation differences between left and right image features. Since we use noisy correspondences to update the epipolar geometry, we also use a simple alternation technique that increases the robustness.

The retinal domain method extends a coarse-to-fine multi-orientation phase-difference stereo disparity algorithm (Sabatini et al., 2010 [J16]). A closely-related optical flow algorithm has been shown to be suitable for real-time implementation on graphics hardware (GPUs) (Pauwels & Van Hulle, 2008). As said above, we propose to compensate for large orientation differences between the left and right images by changing the read-out of the filter responses. Instead of computing phase differences from identically-oriented filters in the left and right images we shift the right filter responses across orientation.

The method corrects an erroneous initial epipolar geometry estimate (e.g. arriving from the motor signal) by applying 3D rotations to the hypothesized left and right camera orientations. This is achieved in practice by warping the images or, more precisely, the filter responses (this avoids the need for re-filtering). The required 3D rotations are derived from vector disparity deviations orthogonal to the currently hypothesized epipolar lines. Vector disparity is similar to optical flow and can be estimated from the phase differences at multiple orientations by allowing each phase difference to constrain the projection of the vector disparity on the filter's orientation. By alternating between geometry and disparity updates in a coarse-to-fine fashion, both steps are performed simultaneously and assist each other.

We can demonstrate the performance of the method on a synthetic example presented in Fig. 20. In this example, we have constructed an unrectified image pair by warping the left and right images of the *venus* stereo pair according to arbitrary 3D rotations of the left and right cameras. An anaglyph of the two images is shown in Fig. 20(A). The left image is in the red channel, and the right image in the green and blue channels. It can be seen that the images differ in terms of horizontal and vertical shifts, rotations and stereo disparity. The ground truth epipolar disparity is shown in Fig. 20(B). Both the standard and proposed stereo algorithms were applied to this image pair using five scales. The results obtained with the standard algorithm are shown in Fig. 20(C). A simple left/right consistency check has been performed here to remove unreliable estimates. If the difference is more than one pixel, the estimate is considered unreliable and removed. The same procedure can be applied to the proposed algorithm. In the right-to-left stage, the estimated geometry can be re-used. The results obtained with the proposed method are shown in Fig. 20(D). The method was initialized assuming a rectified situation. Within each scale, the epipolar geometry estimation typically converges after five updates. Therefore five internal iterations were performed here. Clearly, a much larger number of consistent estimates are found than in Fig. 20(C), and the estimates closely resemble the ground-truth epipolar disparity.

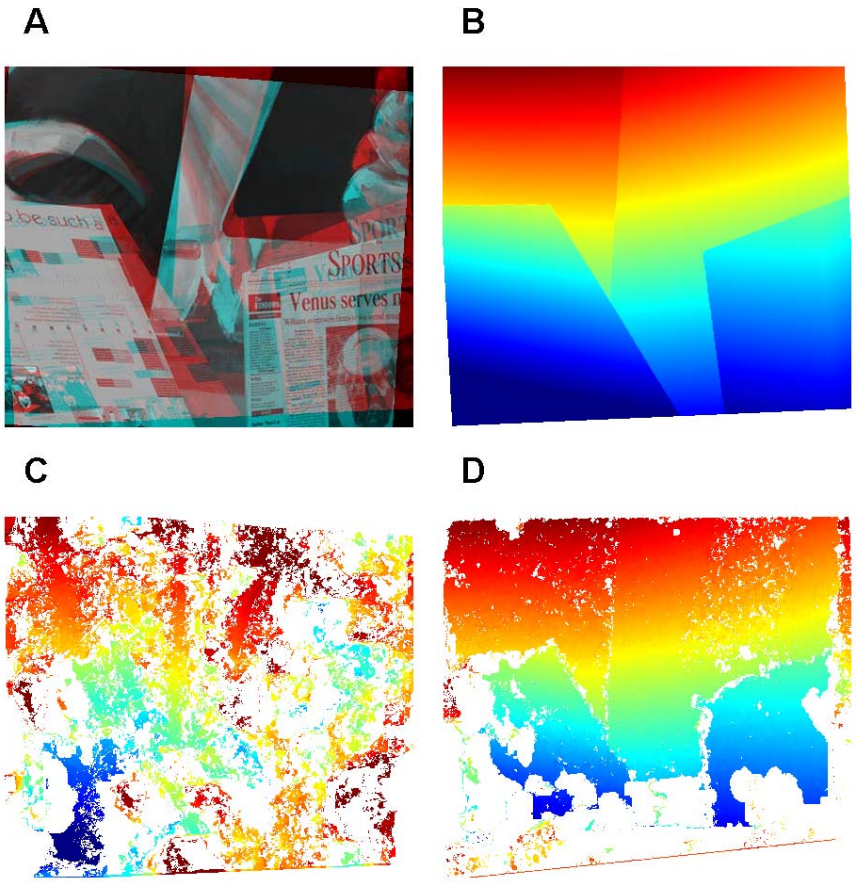


Figure 20: *Performance of the retinal domain method. Ground truth epipolar disparity (B) and consistent estimates obtained with the standard (C) and proposed (D) method on a synthetically generated image pair (A).*

Auto-calibration in the Cortical Domain

We next discuss an extension of the algorithm that also includes a space-variant mapping from the retinal to the cortical domain using the log-polar transformation (Schwartz, 1977). Many mappings from retinal to the cortical domain have been proposed, mainly differing in the way they deal with the singularity in the center. We use the Central blind-spot (CBS) model (Capurro et al., 1997). Design rules for this model can be found in (Traver & Pla, 2008). An example transformation is shown in Fig. 21. The original (retinal) image is shown in Fig. 21(A), and the transformed (cortical) image in Fig. 21(B). The pixel size is the same in both images to show the compression factor achieved. Figure 21(C) contains the retinal image obtained by transforming the cortical image (B) back to the retinal domain using interpolation. This makes it clear how the precision is retained in the fovea (except of course in the blind spot), and gradually reduces towards the periphery.

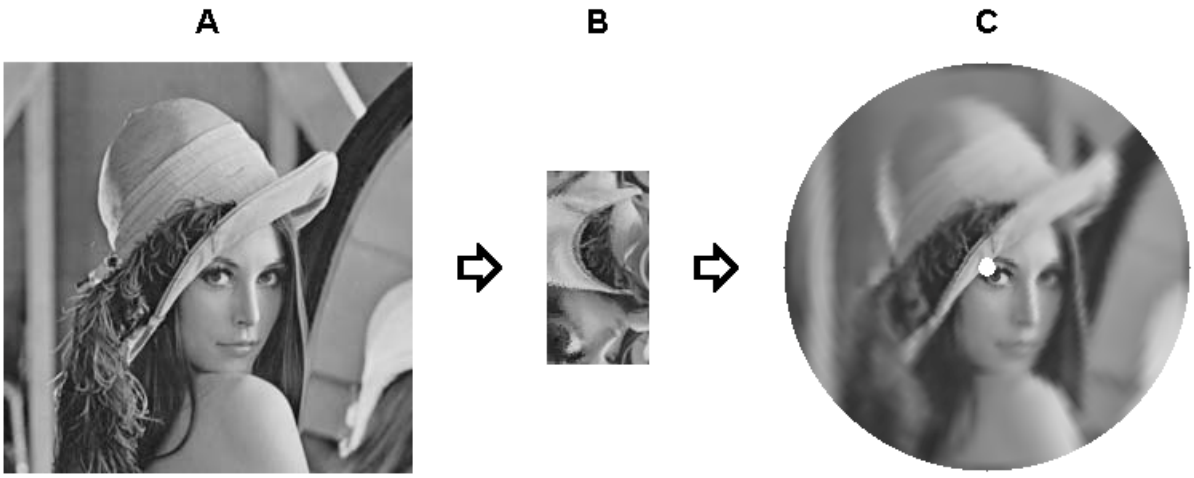


Figure 21: *Example illustrating the transformation of an image from retinal (A) to cortical (B) coordinates using the central blind-spot model. The size of pixels is the same in both images. Image (C) contains the cortical image mapped back to the retinal domain.*

As shown in (Solari et al., submitted [J8]), the same phase-difference technique as used in our retinal-domain method can be applied to estimate the correspondences in the cortical domain with good precision (for more details on the design rules used for the log-polar transformation, see the note at the end of this section). This means applying the same filterbank directly to the cortical domain (which saves considerable computational resources), and estimating the correspondences by integrating the oriented phase differences into a vector disparity estimate. The cortical domain algorithm only differs from the retinal domain algorithm in that the filtering and correspondence estimation are done in the cortical domain. The updates to the epipolar geometry, and the corrections of the estimated vector disparities to the current epipolar geometry estimates are still performed in the retinal domain. This does require intermediate coordinate transformations. To transform vectors between the two domains, the transformation equations are simply applied to the vectors' start- and endpoints. One other difference between the two algorithms is that we do not (yet) compensate for orientation differences in the cortical domain.

We again use the *venus* stereo pair to construct two different scenarios, and examine the improvements in disparity and geometry estimation with the proposed algorithm in the cortical domain. In both scenarios, we use three scales and initialize the geometry to a rectified situation (horizontal epipolar lines).

The first scenario is concerned with a camera setup close to vergence (in the image center), but disturbed by small inaccuracies in the hypothesized camera geometry (rectified). The anaglyph for this situation is shown in Fig. 22(A). Note that the disturbances result in epipolar lines that are clearly not horizontal (see e.g. how the 'V' of Venus has shifted upwards). The disturbances involve 3D rotations along all axes. The magnitude of the retinal ground truth vector disparity is shown in Fig. 22(B), which confirms that the center is close to vergence. The cortical image pair is shown in Fig. 22(C,D), with cortical ground truth vector disparity in Fig. 22(E). The results obtained with and without the proposed auto-calibration technique are shown in Fig. 22(F) and (G) respectively. Figure 22(G) was obtained by computing disparity along the epipolar lines directly in the cortical domain using a coarse-to-fine refinement procedure (similar to an optical flow algorithm). Note how the auto-calibration greatly improves the results and more closely resembles the ground-truth, particularly in the outer regions of the fovea (the central columns in Fig. 22C-G). The improved estimates there can then be used to further refine the vergence. The mean and standard deviation of the errors (vector differences w.r.t. ground-truth) for the two methods can be found in Table 3 and confirm this improvement.

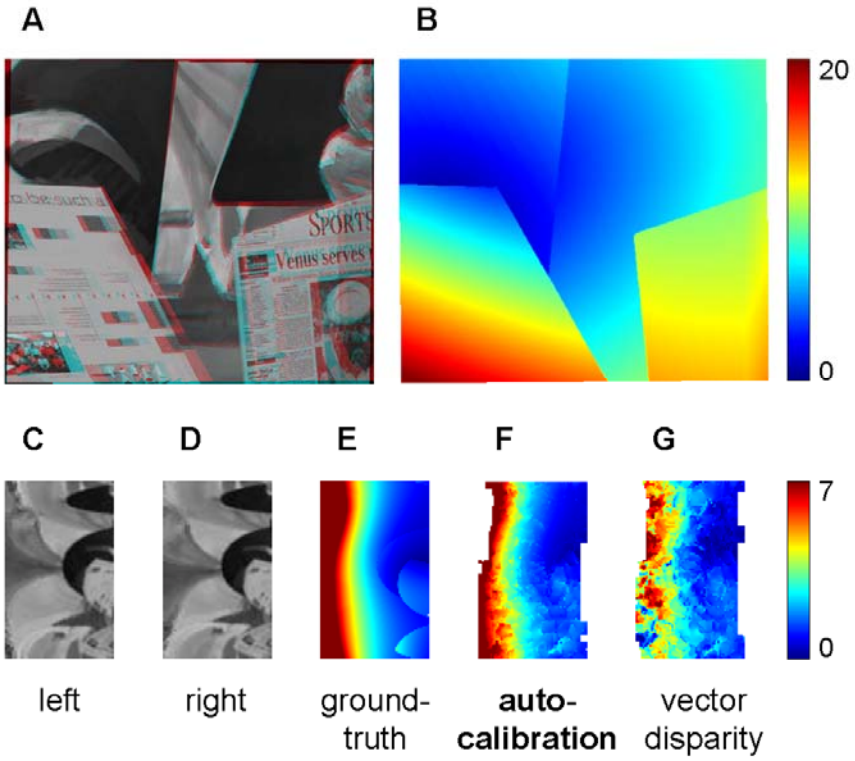


Figure 11: *Improvements of auto-calibration over vector disparity in a scenario close to vergence with small errors in the geometry estimates. (A) Left and right retinal image anaglyph showing non-horizontal epipolar lines. (B) Magnitude of retinal ground truth vector disparity. (C) Left and (D) right cortical images. (E) Magnitude of cortical ground truth vector disparity. Magnitude of cortical disparity estimated along the epipolar line with (F) and without (G) auto-calibration.*

	auto-calibration		vector disparity	
scenario	mean	std	mean	std
close to vergence	1.03	1.40	1.70	2.44
away from vergence	3.43	4.28	10.72	10.20

Table 3: *Mean and standard deviation (in pixels) of the magnitude of the vector differences between ground truth and estimated vector disparity.*

We also examined a very different scenario, far away from vergence and with large errors in the hypothesized geometry (again rectified). The anaglyph for this situation is shown in Fig. 23(A). It is clear from the magnitude of the retinal ground truth vector disparity in Fig. 23(B) and the cortical images in Fig. 23(C,D) that the fovea contains very different image parts and that the vector disparities are much greater everywhere (and again clearly not horizontal as in a rectified situation) than in the previous scenario. As before, the results obtained with and without the proposed auto-calibration technique are shown in Fig. 23(F) and (G) respectively. Again we see that the proposed method greatly improves the results. This is also measured quantitatively in Table 3. We did use the prior knowledge about the lack of vergence in this situation, and did not use orthogonal disturbances in the fovea to update the geometry. Instead we only relied on peripheral errors here. Using foveal errors worsened the results but did not make the algorithm fail.

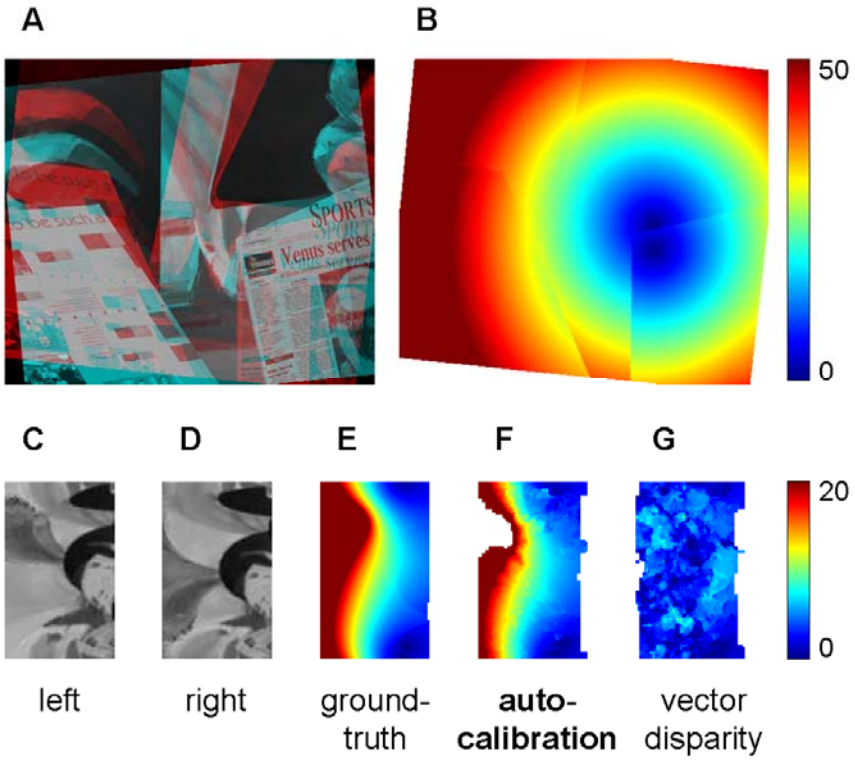


Figure 23: *Improvements of auto-calibration over vector disparity in a scenario away from vergence with large errors in the geometry estimates. (A) Left and right retinal image anaglyph showing non-horizontal epipolar lines. (B) Magnitude of retinal ground truth vector disparity. (C) Left and (D) right cortical images. (E) Magnitude of cortical ground truth vector disparity. Magnitude of cortical disparity estimated along the epipolar line with (F) and without (G) auto-calibration.*

To demonstrate that the algorithm is capable of making large corrections to a hypothesized camera geometry, we also show the recovered epipolar geometry for the second scenario in Fig. 24. This figure shows corresponding epipolar lines for a few selected key points. The blue points in the left image (A) correspond to the epipolar lines in the right image (B) and vice versa. Note that very large changes have been made considering that the algorithm started from a rectified situation, and that the epipolar lines are quite accurate (even in the fovea).

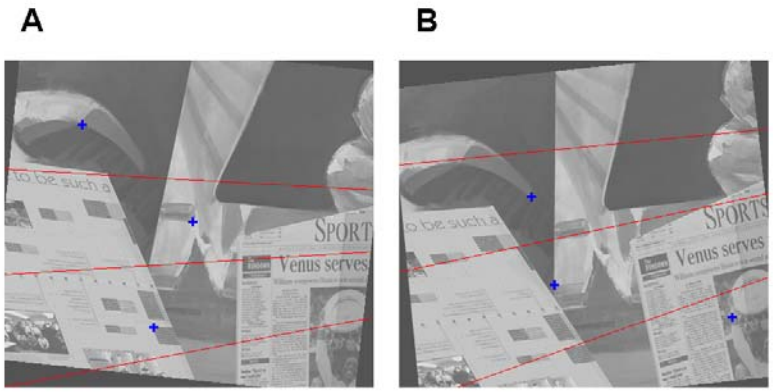


Figure 24: *Recovered epipolar geometry for the scenario of Figure 4. Red epipolar lines in the right image (B) correspond to the blue keypoints in the left image (A) and vice-versa.*

Future Extensions

The next steps will first require a modification of the procedures to the population-based methods for disparity estimation (Chessa et al., 2009a). These are conceptually very similar to the phase-difference approach used here, but allow the application of very specific (gain) modulation procedures. These gain modulation mechanisms will be used to transform the responses from eye- to head-centered coordinates. In addition, the population responses also allow for a closer interaction with the vergence mechanisms developed in Task 2.1.

Note on Log-polar mapping

In the literature, several log-polar mapping models are described (Bolduc & Levine, 1998; Jurie, 1999; Florack, 2007). We have chosen the central blind-spot model (Traver and Pla, 2008).
UG performed a systematic analysis to see how the energy ratio between the response E_{mapped} of a mapped filter (i.e., a filter mapped into the cortical domain) and the response $E_{matched}$ of a matched filter (i.e., a filter directly defined in the cortical domain), is affected by the relationships between the parameters of the log-polar mapping and of the Gabor filter (see Fig. 25). If the aspect ratio of the log-polar pixel is approximately 1, the energy ratio $E_{mapped}/E_{matched}$ remains high, independently of the eccentricity in the cortical plane and of the orientation of the Gabor filter. Conversely, values of the aspect ratio different from 1 yield to lower responses of the filters with respect to the eccentricity and to an anisotropy with respect to
the orientation of the filter. Moreover, Fig. 25 shows that the maximum response is obtained when the spatial support of the filter is small (e.g. 11x11 pixels). It is worth noting that under these conditions the deformations of the mapped filters are relatively small (see inset A of Figure 25a). Once fixed the aspect ratio of the log-polar pixel equals to 1, the influence of the spatial support of the filter can be evidenced from Figure 25b. The response of the filters decreases with an increase of the spatial support, independently of the eccentricity in the cortical plane and of the orientation of the filter. This could be also evidenced looking at the deformed profiles of the Gabor filters (see Figure 25b, profiles marked by D and C). For a given value of the spatial support (e.g. 11x11 pixels) the responses of the filters neither depend on the eccentricity in the cortical plane nor on the orientation of the filter.

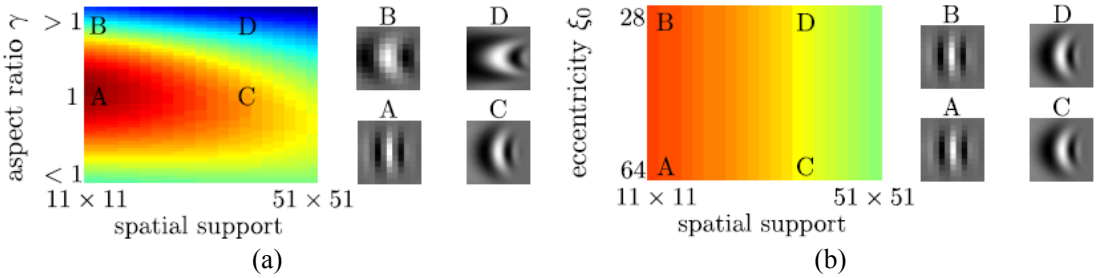


Figure 25: *Variation of the energy ratio $E_{mapped}/E_{matched}$ with respect to the parameters of the log-polar mapping and of the Gabor filters (left side of each subfigure) and profiles of the mapped filters for particular choice of such parameters, marked by capital letters A-D (right side of each subfigure). Hot colors mean high energy ratios, whereas cold colors mean low energy ratios. (a) Aspect ratio of the log-polar pixel with respect to the spatial support of the Gabor filters. The maximum energy ratio is obtained for a squared log-polar pixel and a small spatial support (11x11 pixels). It is worth noting that under these conditions the filters show no deformation (see A), otherwise high deformations are present (see B-C-D). (b) Eccentricity in the cortical plane with respect to the spatial support of the filter. The maximum energy is obtained for a small spatial support.*

This analysis allows us to devise the constraints for the parameters of the log-polar mapping and of the spatial filters, in order to obtain a signal processing in the cortical domain equivalent to the one in the Cartesian domain. Thus, it is possible to perform a direct visual processing in the cortical domain, without adapting the algorithm developed for the Cartesian domain.

Disparity computation in log-polar images

The distributed algorithm for the computation of the 2D vector disparity (Chessa et al., 2009a) is suitable to be directly applied on cortical images, since 2D vector disparity is computed without an explicit search of the correspondences, between the left and the right images, along the epipolar lines. In this way, it is not necessary to take into account that the straight lines in the Cartesian domain become curves in the log-polar space (Schindler, 2006).

In order to quantitatively benchmark the described approach, stereo sets with available ground truth disparities are necessary. To this aim, the tool described in Chessa et al., 2009b is used. Such a tool, exploiting the ground truth available from a 3D model of the observed scene, virtual or real, and the related projected stereo images, provides a way to validate the behavior of an active vision system in a controlled and realistic scenario. Figure 26 shows the left image of a stereo pair used in this analysis, the ground truth horizontal and vertical disparity maps, and the computed disparity maps.

.

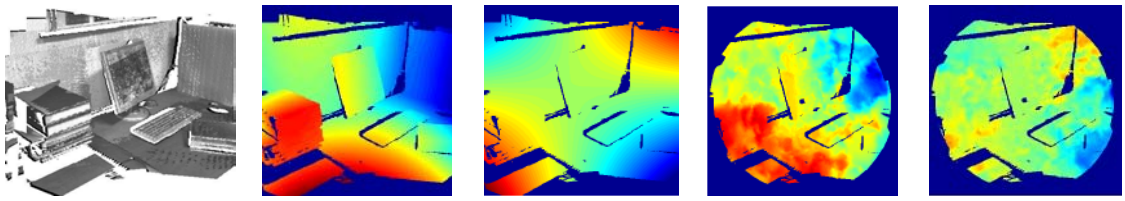


Figure 26: *Comparison between the estimated disparity maps and the ground truth for a stereo pair obtained from a real scenario acquired by a laser scanner. The average errors in the computation of the horizontal (HD) and vertical (VD) disparities are 1.50 and 0.57 pixels, respectively. The ground truth disparity range is between -13 and 21 pixels.*

Furthermore, the reliability of the disparity values with respect to the parameters of the mapping is analyzed. Table 4 shows how the size of the cortical image and the number of the considered scales affect both the execution time, and the global average error on the computation of the disparities, with respect to the ground truth.

In addition to the global average error, computed by considering all the pixels of the image, the average error around the fovea (a region with a radius half of the image size) and in the periphery is computed separately. This approach is necessary, since the central part of the image is of great importance for active vision tasks and the error in the peripheral area is affected by the increased size of the log-polar pixel. The analysis show that the average error in the region around the fovea is small, i.e. less than 1 pixel in every condition. The execution time is expressed as a fraction of the algorithm execution time in the Cartesian domain with the optimum set of parameters, in this way, the obtained results are not bound to a specific implementation. It is worth noting that the time necessary for the forward and backward log-polar transformation is a small percentage of the total execution time.

CARTESIAN DOMAIN								
Image size	Number of scales	AEH	AEV	AEH fovea	AEV fovea	AEH periphery	AEV periphery	Execution time
331x331	5	0.82	0.27	0.10	0.10	0.83	0.28	100%
331x331	2	3.50	2.43	0.30	0.21	3.59	2.68	89%
331x331	1	3.73	2.74	0.76	0.54	3.77	3.09	67%
CORTICAL DOMAIN								
Cortical image size	Number of scales	AEH	AEV	AEH fovea	AEV fovea	AEH periphery	AEV periphery	Execution time
100x159	2	1.37	0.45	0.29	0.29	1.53	0.61	36%
100x184	2	1.31	0.39	0.28	0.22	1.47	0.57	36%
100x159	1	1.92	0.64	0.43	0.35	2.09	0.83	29%
100x184	1	1.95	0.71	0.46	0.29	2.07	1.01	20%
64x117	1	2.13	0.84	0.50	0.34	2.23	1.04	9%

Table 4: *Performance comparison between the computation of the disparity in the Cartesian and in the log-polar domain for different sizes of the cortical image. The global average error for the horizontal disparity (AEH) and for the vertical disparity (AEV), together with the local error around the fovea (AEH fovea and AEV fovea) and in the periphery (AEH periphery and AEV periphery) are shown. The execution time is expressed as a percentage of the execution time in the Cartesian domain with three spatial scale.*

Deviations from the project workprogramme

None. As an extension to the work planned in Annex I, a specific analysis on the joint design on the joint design constraints for the discrete log-polar mapping and for the bank of Gabor filters became necessary to guarantee minimal distortions of the mapped filters across retinal eccentricity for a proper multi-orientation and multi-scale processing of the visual signal. A MATLAB module implementing the resulting optimized mapping has been developed and made available to the Consortium.

WP3: Selecting and binding visual fragments

Leader: Fred Hamker (WWU)
Contributors and planned/actual effort (PMs) per participant:
WWU (23/23) and UG (3/2)
Planned/actual Starting date: Month 1/1

Workpackage objectives

This workpackage is devoted to develop novel concepts of selecting and binding within a fragmented 3D scene representation. One of those fragments is object identity. Object identity will be obtained from a bi-directional, hierarchical representation of learned feature detectors. The development of the appropriate learning rules will be an essential part of this project, since the learned connections will be used for the selection of a fragment. In the first period we aimed at learning V1-like feature detectors form stereo images. Beyond object identity, a distributed representation requires to actively bind and represent the relevant visual fragments for the task at hand. Thus, we study how attentional dynamics allow to actively bind features and build task relevant representations. Moreover, we will develop a novel framework for the task relevant binding of fragments in a global workspace using reward-based learning.
Starting point has been a revised model for learning V1 receptive fields (Wiltschut & Hamker, 2009) and models of attentional dynamics (Hamker, 2005, Hamker & Zirnsak, 2006, Hamker et al., 2008) and an overview paper about the role of the Basal Ganglia in cognitive control (Vitay et al., 2009 [J9]).

Progress towards objectives

Task 3.1 is devoted to develop novel concepts of learning to encode a fragmented 3D scene representation.
The human brain actively generates a cognitive interpretation of a perceived scene. It doesn't save the scene as a pure 2D image or reconstruct a 3D model. Rather, it creates a very efficient code in which a scene consists of distributed and loose features. Primates use information of both eyes to improve object recognition in a scene. We here describe our results of learning disparity and object selective cells in the visual pathway. Previous models of primary visual cortex (V1) encoding disparity have primarily been constructed by hand to fit specific data.
As described in the previous report a year ago we have developed a bi-directional network with non-linear dynamics based on Hebbian and anti Hebbian learning principles where cells, after having learned from natural stereo image scenes, develop disparity- and feature-selective properties comparable to cell characteristics measured in V1. In contrast to constructed disparity encoding cells that are commonly used to model binocular vision in V1 and beyond, our learned cells show more complex tuning characteristics (Fig.27) of which the categorisation introduced by Poggio et al. (Poggio 1995, Poggio & Fisher 1977, Poggio et al. 1988, Poggio & Talbot 1981) is only a subset of. The disparity tuning characteristics of the learned cells probed with displaced random dot images (RDI) exhibit complex patterns with regard to horizontal and vertical displacement of the input. Reduced to zero vertical displacement exemplary cells for the categorisation of disparity tuning characteristics introduced by Poggio and collaborators were found (Fig. 28) but overall there was no clustering of cells with specific properties to support such a distinction into classes of disparity selective cells.

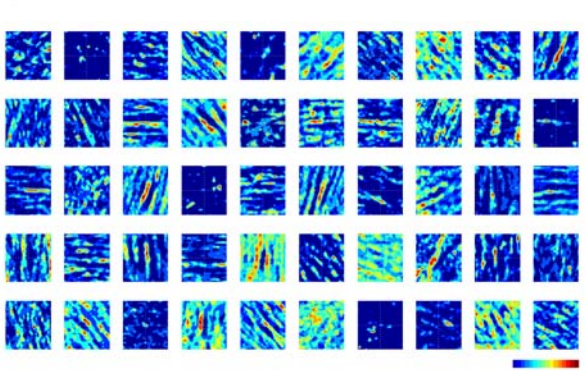


Figure 27: *Exemplary disparity tuning plots of the learned V1 like cells probed by displaced random dot images.*

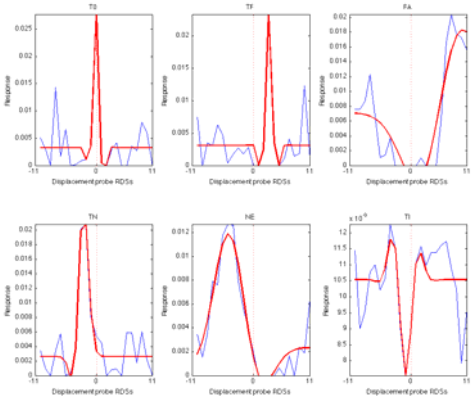


Figure 28: *Exemplary disparity tuning curves according to the categorisation introduced by Poggio et al..*

With respect to ocular dominance the Gabor fitted learned receptive fields (RF) exhibit also varying influence of the two eyes, from monocular to binocular response properties, indicated by the amplitudes of the fits in the two eyes as shown in Figure 29.

There is a strong correlation between the orientations of the Gabor fitted RFs in the two eyes as they only differ by a small magnitude in most cases as one can see in Figure 30. The results are in line with data from the cat (Blakemore et al. 1972, Nelson et al. 1977) and also with data from macaque V1 neurons (Bridge and Cumming 2001) and all of these suggest that binocular neurons in the early visual system have similar orientations in both receptive fields and therefore have no separate encoding scheme for orientation disparities.

Barlow et al. (1967) have found an orientation anisotropy for horizontal and vertical disparities as they measured a larger range for horizontal (±3.3°) than for vertical disparities (±1.1°). In contrast to that, subsequent studies (Ferster 1981, Joshua & Bishop 1970, LeVay & Voigt 1988, Nikara et al. 1968, von der Heydt et al. 1978) did not show such an anisotropy for small eccentricities. The results of our learned receptive fields also exhibit little differences of the encoding range for binocular disparities between horizontal and vertical disparities.

The learned receptive fields in our model show both phase and position-based mechanisms to encode binocular disparity. Our results suggest that there is no categorization into distinct classes of disparity encoding neurons into phase or position encoding neurons but that both mechanisms encode disparity in V1 at the same time with smoothly varying influence over the whole population. Even though it is not clear so far what is the contribution of these two mechanisms to binocular vision other reports (Prince et al. 2002, Anzai et al. 1997, Anzai et al. 1999) agree with the view that both phase- and position-based mechanisms are used to encode disparities and their contribution to binocular vision was found to be very similar.

As positional differences parallel to the orientation of the receptive fields cannot be detected, also known as the aperture problem, many artificially constructed disparity detectors assume that disparities are encoded in the direction orthogonal to the orientation of their receptive fields. To compare our results with this model we have calculated the angle between the orientation of the sinusoid of the Gabor fits and the corresponding positional difference vector. Our results are not in compliance with the perpendicular model of the relation between disparity direction and receptive field orientation as there seems to be no correlation between these two parameters.

In summary we have developed a model of V1 that is capable of unsupervised learning of disparity-tuned feature selective cells from a set of stereo image scenes. The learned cells are in compliance with physiological data in many accounts but show tuning characteristics that are much more complex than most

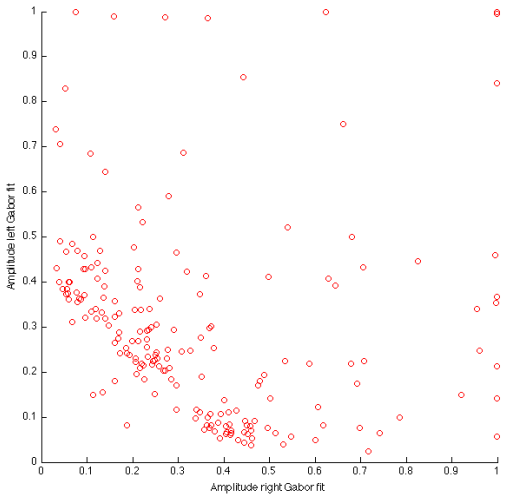


Figure 29: *Ocular dominance indicated by the amplitude of the Gabor fitted RFs in the two eyes.*

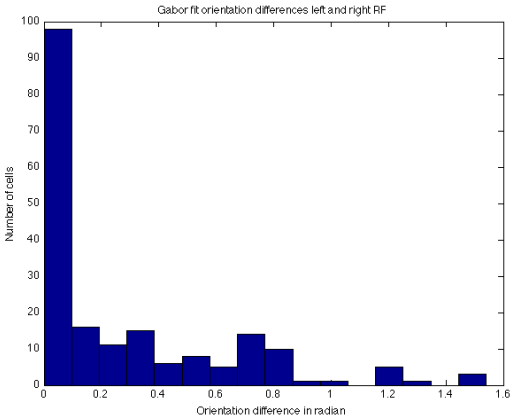


Figure 30: *Histogram of the orientation differences of the Gabor fitted RFs in the two eyes.*

constructed binocular units used in modelling V1 so far. The possible influence of these differences and how it might contribute to binocular vision, depth perception and object recognition must be the scope of further research.

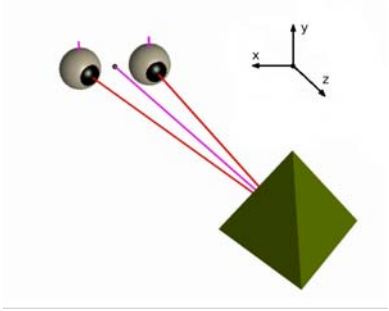


Figure 31: *The coordinate system and the arrangement of the eyes, object and fixation position in the raytracer.*

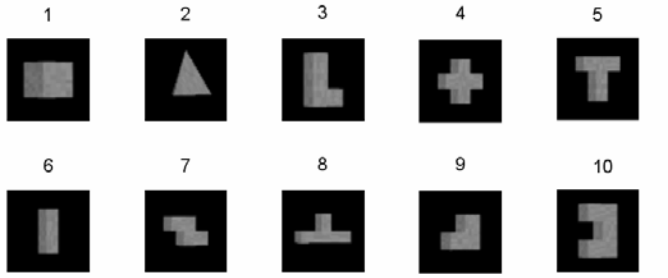


Figure 32: *The stimuli consist of 10 different 3D objects. Here we show the image of the left eye.*

The second subtask focuses on the extension of the learning process to higher cortical areas (in this case the areas V2 / V4). The goal is to use biological motivated learning mechanisms to obtain objective selective cells with intermediate complexity that show disparity tuning characteristics (see D3.1b).
Disparity in a stereoscopic stimulus can result from two different effects. The object position can be shifted (while fixation is constant) or the fixation point can be shifted (while object position is constant). We tested the cross combination of both effects. Each object is presented for a probabilistic time period and the network responses determine the adaptation of the weights via a trace learning paradigm. We did not use a supervised learning paradigm because of its biologically implausibility.
The input stimuli are the left and right eye view of a 3D object. We chose a raytracer engine (developed by Chumerin in WP2) to produce the images (Fig. 31) and we compiled 3D models of cubes to create 10 different objects (Fig. 32). We used a rate coded neuron model to describe the neural network which consists of three layers. The first and second layer process both images to detect disparities via an energy model

(Chessa et al., 2009a). The responses serve as input in the third layer which represents object and disparity selective cells. We used a general learning algorithm (Wiltschut & Hamker, 2009) which is motivated by biological research to capture the basic principles of primate 3D perception and modified it with a trace in the output cells. Thus, the neural response of the previous view is linked with the present view. As a result the response of the cells will be invariant to the changes at a small temporal scale in the sequence. It uses post-synaptic inhibition, Hebbian Learning and is able to learn simultaneously feed forward (stimuli driven) and feedback (attention driven) connections.

Two problems should be solved with this model. The first one is the learning and recognition of objects and the second one is to show disparity selectivity in the cell responses. Our results show that regardless of the number of different objects and independently of the number of cells (as long that there is at least one cell per object) the model is able to learn and discriminate (Fig. 34) all objects. Figure 33 shows the average response of each cell to the different objects. It can be seen that each object is learned by several cells and thus an object is characterized by a specific population code.

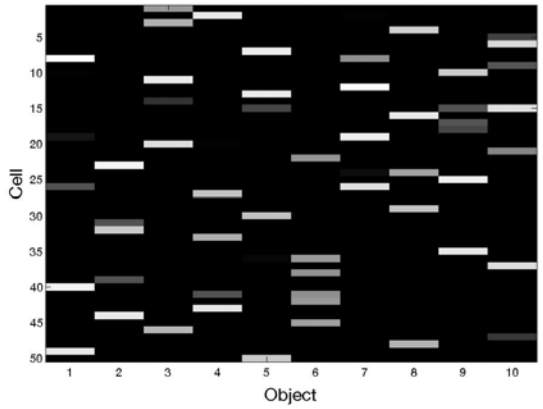


Figure 33: *The selectivity of each cell for the objects. For each object (x-axis) the average firing response of each cell (y-axis) is plotted. The strength of the average firing pattern to an object is characterized by the brightness (0 dark, 1 bright). Every object is represented by a combination of different cells with nearly no overlap to other objects.*

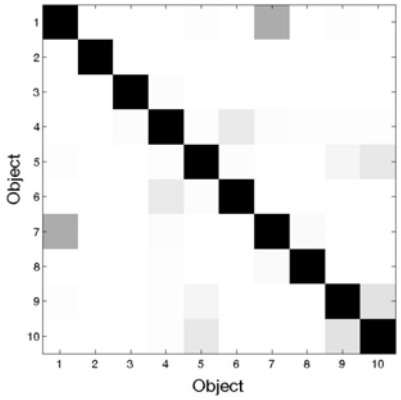


Figure 34: *Discrimination between the objects. Using the discrimination value (see D3.1b), the similarity of the average response (shown in Fig. 33) to an object is shown here. The brightness represents the dissimilarity of the population code responding to the objects.*

Although each object is coded by several cells, there are three kinds of typical disparity tuning behaviors:

- Type 1 cell: The cell is tuned to a specific disparity in relation to the fixation point. Thus it responds best to specific combinations of object distance and fixation point. This also includes a certain degree of pooling because the cell responds only to a specific absolute disparity which is invariant in object position along the z-axis.
- Type 2 cell: The cell is tuned to an object and is completely invariant to object position or fixation point position. This cell performs object recognition in an optimal way, because its firing rate is independent of all but the information provided by the object itself. This might indicate that the cell is tuned to the edge information with contribution of the object depth information.
- Type 3 cell: This type is a mix of the first two types. It shows good object recognition (as Type 2) but with a slight preference to a specific absolute disparity to the fixation point (as Type 1).

Additionally we investigated the complexity of the object selective cells. The degree of invariance is a core idea of the distinction between simple and complex cells in the brain. The receptive field of a simple cell consists of spatially distinct excitatory and inhibitory regions whereas a complex cell combines these regions (De Valois et al. 1982). Thus, the object recognition cells in the third layer loosely correspondent to V2 or V4 cells. All of them are spatially invariant with different degrees of disparity invariance.

To summarize, we have developed a model that learns with self-organized cells and Hebbian learning from simple edge and disparity cell responses an object recognition representation of intermediate complexity that shows tuning characteristics in absolute disparity. The degree of the influence of object depth for object recognition will be the focus of future research.

Task 3.2: ***Selecting visual fragment***

This task aims at understanding and developing mechanisms of attentional selection. Before we apply our developed mechanisms to scenes containing objects we first describe their relation to biological data.

Our previous model (Hamker, 2007) predicts feature-based distortions of the population responses. The general idea of this model is that the gain resulting from feature-based attention is not uniform across the whole population, but rather peaks around the attended feature, which in turn distorts the population response such that it is effectively shifted towards the attended feature. We recently obtained data from own psychophysical experiments that allowed us to test if this model prediction is meaningful. Thus, we here describe our effort to fit our model to this experimental data. We observed that direction estimates of the static motion aftereffect (SMAE) drastically change when human observers attend to a stimulus whose motion direction differs from the one of the adaptor. This observation can be likely generalized to other features such as orientation. It suggests that feature-based attention might operate by local magnifications of feature space between relevant and irrelevant features. The influences of attention were simulated in explicit terms. Tuning curves are described by Gaussian functions (Fig. 35A). The influence of attention on the population response is formalized as a difference of Gaussians, that is, neurons which prefer directions close to the attended target are enhanced while neurons which prefer directions farther away are suppressed (Fig.35B). A simple Gaussian did not allow to fit the model to the data. Please refer to Zirnsak & Hamker (in press) for details about the model. Consistent with electrophysiological recordings (Martinez-Trujillo and Treue, 2004) this results in a sharpening of the population response if target and adaptor have the same direction (Fig. 35C) as determined by the half-maximum width of the response profile. For adaptors close to the target, however, the gain modulation leads to a distortion of the population in such a way that the population vector is shifted towards the target direction (Fig. 35D). Similarly, the vector of the population response to stimuli which lie in the suppressive surround is shifted away from the target direction. Altogether the distortions lead to local magnifications in feature space consistent with the response of the human subjects (Fig. 35E). While these effects already occur due to simple gain changes of single neurons without shifts in the tuning curve, neurons that compute a weighted sum of the distorted population response ought to also alter their tuning properties, as predicted by the model (Fig. 35F). Tuning curves which are located in feature space close to the center of attention are attracted, whereas others located farther away are repelled. However, due to the suppressive effects surrounding the center of attention an effective increase in processing occurs at the attended direction. Thus, the whole population preferably processes the attended feature. This is the first time that local distortions of feature space have been predicted and experimentally verified. For details please refer to Zirnsak & Hamker (in press) [J13].

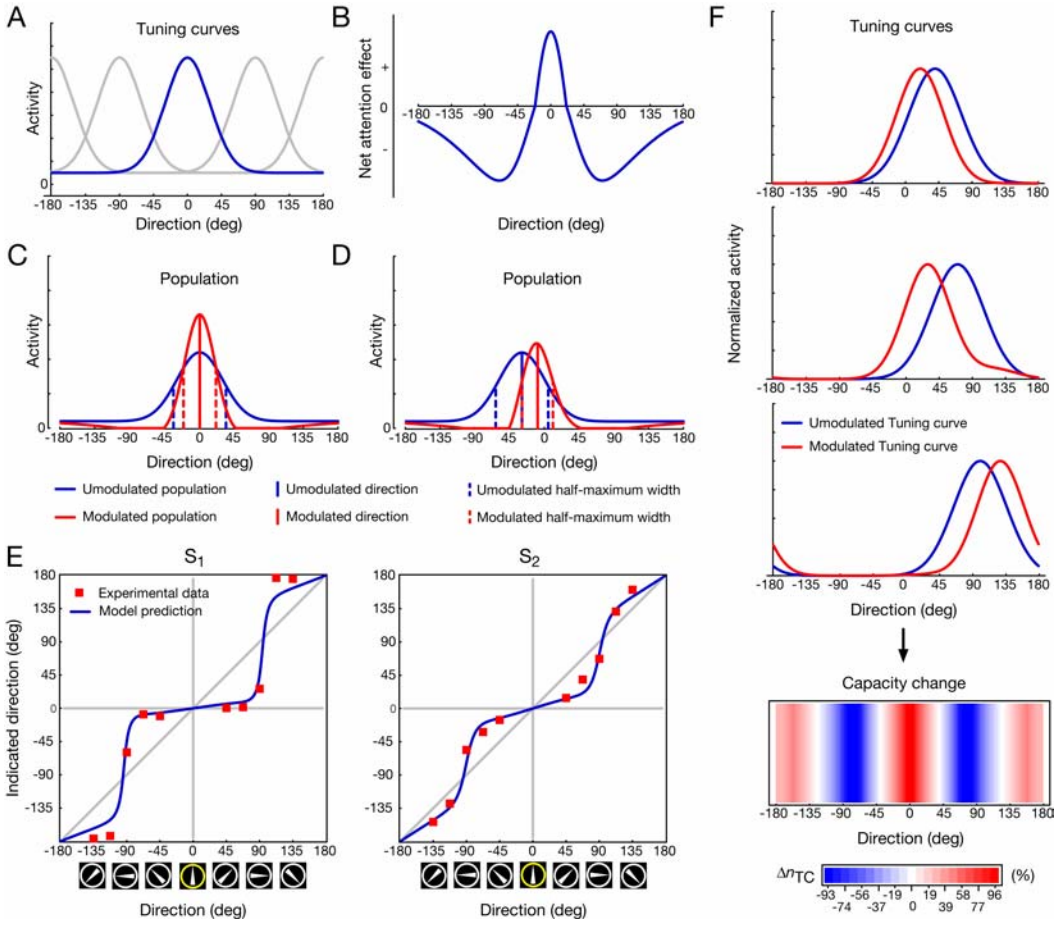


Figure 35: (*A), Tuning curves of sensory neurons located on a circle in arbitrary units (a.u.). (B), The net effect of attention. Directions close to the attended direction are enhanced while directions farther away are suppressed. (C), Consistent with experimental data the population response is sharpened as indicated by the half maximum response when the stimulus direction is equal to the attended direction. (D), For a stimulus with a direction of -30° the population response is distorted, that is, the population vector of the neural activity is attracted towards the attended direction. (E), Indicated SMAEs of subject 1 and 2 for ten directions denoted by the red squares shown together with the model fits (blue line) as a function of the veridical direction. Note again that SMAEs are shown as the direction which would have caused them assuming a direct reversal. While directions near the attended direction 0 are attracted, directions farther away are repelled. (F), Exemplary tuning curve shifts of three model neurons. The center of the unmodulated tuning curve is 40°, 70°, and 100° from top to bottom. While the center of the tuning curve shifts towards the attended feature for tuning curves with unmodulated centers close to the attended direction, it is shifted away for tuning curves with unmodulated centers farther away. This results in an increase of tuning curves which are processing the attended and the opposite direction. However, since the tuning curves with preferred direction farther away from the attended direction are suppressed, an effective increase in processing occurs at the attended direction, that is, tuning curves of neurons shift their preferred direction such that the whole population preferentially processes the attended properties.*

After having shown that the model relates well to psychophysical data we now describe our effort to apply these concepts to real images. We simulate two areas, an earlier area (V1) and a higher area containing object selective cells (similar to V2/V4). As input stimuli we used the left and right eye view of a 3D object. We use the raytracer engine (developed in WP2) to create 3D models composed of cubes and to produce the stereo images. The first model layer relates to area V1 and processes both images to detect disparities via an energy model (Sabatini et al., 2007). The responses serve as input in the second layer (V2/V4) which

represents object and disparity selective cells as described in Task 3.1. We used a weight sharing approach to analyze the whole visual scene in parallel, i.e. the detection of objects is independent from the location of the object in the visual scene on a 2D plane. A top-down "Attention signal" can bias a particular object for selection (Fig. 36). An oculomotor loop via the frontal eye field (FEF) can select the location of a particular object for a saccadic eye movement and also provides a spatially selective attention signal.

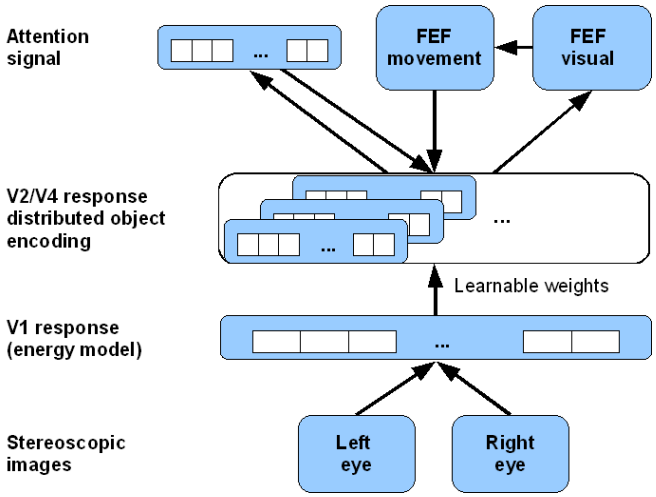


Figure 36: *The neuronal network architecture of the model. The V1 layer combines the stereo images and the layer V2/V4 has been trained to represent simple object shapes.*

The binding process to select a visual fragment operates continuously, but it can roughly be illustrated by two processes. One operates in parallel over all fragments and increases the conspicuity of those that are relevant for the task at hand, independent of their location in the visual scene. The attention signal stores the features representing the task relevant visual fragment and reinforces them in V2/V4. The other is linked to action plans, here the eye movement, and binds those fragments together, which are consistent with the action plan, typically by their location in the visual scene. The loop over the FEF visual and movement maps realizes this idea. Both processes use competition to decrease the activity of irrelevant features and locations in V2/V4.

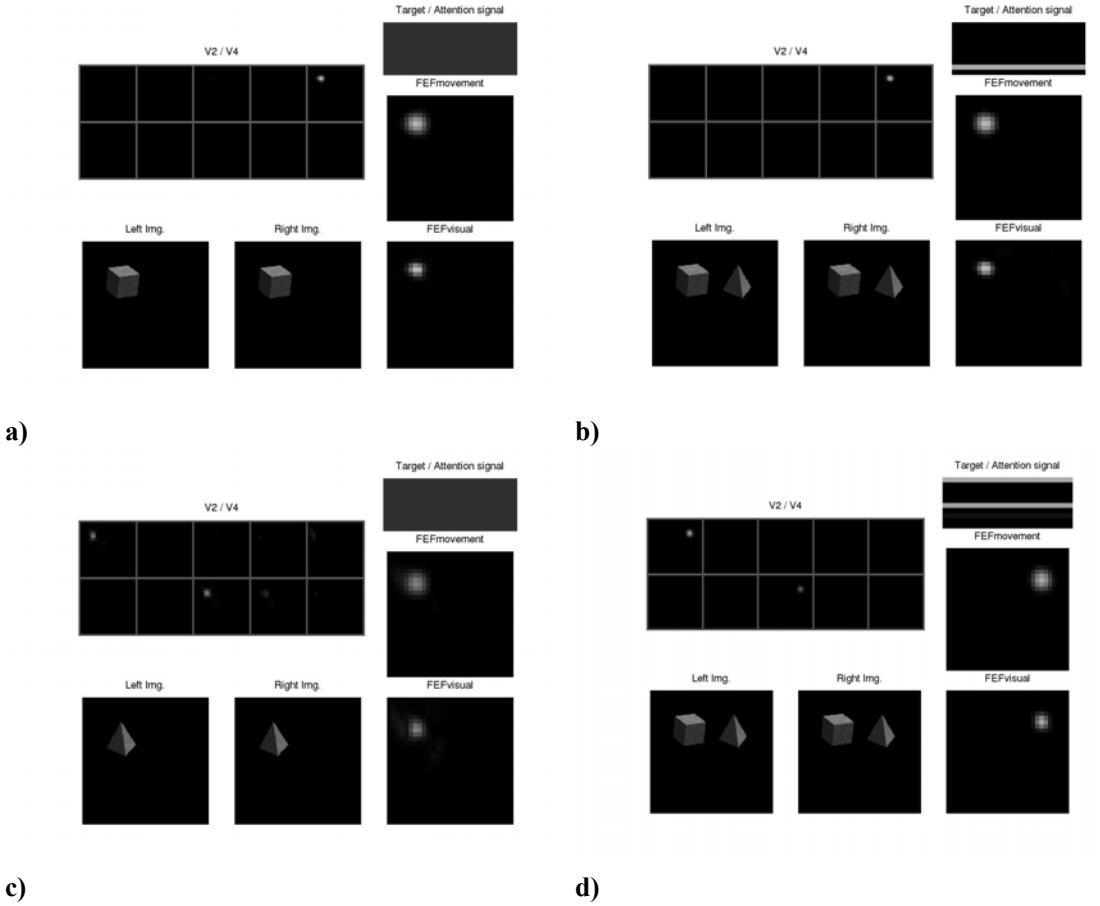


Figure 37: *The figure shows the layer activities during the object localization experiment. The left and right images are presented to the model and the responses of the population code (10 cells encodig features) in V2/V4 are computed as described in Fig.36. Thus each box shows the activity of a single feature in image coordinates. The graphs also show the attention signal (population code on the y-axis) and the two FEF maps. Normally, the x and y axis correspondent to the spatial x and y axis of the images.* ***a)*** *The system searches for the most conspicuous object in the scene, the cube, and stores the V2/V4 response as an attention signal for b).* ***b)*** *The attention signal reinforces the features which represent the cube and the system selects it for an eye movement.* ***c)*** *The system memorizes the tetrahedron and creates an attention signal for the tetrahedron.* ***d)*** *Now the system searches for the other object (tetrahedron) and detects it.*

Fig. 37. demonstrates the binding and the selecting of visual fragments within an object localization experiment:

1. We present first an object alone in a scene without an attention signal (Fig. 37a). The model selects the most conspicuous region (the object) and binds the V2/V4 activation to the working memory, which stores the attention signal.
2. We present a black screen to deplete all cell activities in the system.
3. We test the ability to select visual fragments (an object). We present a cluttered scene (Fig. 37b) (here for simplicity with only 2 objects). The attention signal encodes the visual fragment and reinforces all features in the population code of V2/V4. The system is able to locate the object which was stored in step 1.

In order to show that there is no significant bias in the system to select only a particular object, we repeated the experiment with the other object, the tetrahedron, as the object of interest. After binding the visual fragment (Fig. 37c) the model locates it in step 3 (Fig. 37d). To summarize, we have shown how attention can bias the selection of a particular visual fragment.

Task 3.3: ***Selecting between behavioral alternatives.***

Vision requires high-level cognitive control in form of visual-visual and visual-reward associations, specifically when vision is embedded into a task that requires to interact with the environment. In our developed model, the representation of fragments are coordinated by the thalamus and basal ganglia in order to resolve conflict between competing systems and to schedule tasks in time related to the idea of a central selection device. While the cortex being more involved in learning specific correlations and exact computations, the basal ganglia with the thalamus coordinates these activations by activating specific loops in time.

This task coordination in the basal ganglia can be learned based on stimulus-reward associations. Almost every task requires to hold previously visible information or the recall of memory as context information, but especially when the sharing of workspace is required. Such properties of cognitive vision systems are often referred to establishing a global workspace for specific tasks such as planning and action control. It has been generally suggested that this persistent activity requires the release of dopamine. However, the dopamine system is not specific enough to schedule events in time. We propose that the dopamine system is an arousal system that induces competition and supports short-term memory of salient stimuli by increasing the efficiency of recurrent interactions, but over time the basal ganglia will learn to schedule a task by activating specific thalamocortical loops.

We now describe the model using a generic simple task with artificial input: a delayed match to sample (DMS) and delayed pair association (DPA) task. The model is not limited to these tasks and could in principle be used in any task that requires to link an input to a decision constrained by reward. However, the present version does not solve all aspects of reinforcement learning such as the temporal credit assignment.

The architecture of the model is depicted in Fig. 38(A). Visual inputs are temporally represented in the perirhinal cortex (PRh), each cell firing for a particular visual object. These perirhinal representations project to the prefrontal cortex (dlPFC) where they are actively maintained for the duration of the task. These sustained activations in dlPFC are artificially controlled by a set of gating signals, leaving unaddressed the temporal credit assignment problem. PRh and dlPFC both project extensively to the caudate nucleus (CN), which learns to represent them in an efficient manner according to the task requirements. Depending on reward delivery in the timecourse of learning, each active striatal cell learns to integrate perirhinal and prefrontal information in a competitive manner due to inhibitory lateral connections. This mechanism leads to the formation of clusters through learning in striatal cells that represent particular combinations of cortical information depending on their association to reward. These CN cells send inhibitory projections to the SNr, whose cells are tonically active and learn to become selective for specific striatal patterns. This learning between CN and SNr is also dependent on reward delivery. Learning of the lateral connections between SNr cells additionnaly allows to limit the number of simultaneously inhibited SNr cells. These cells in SNr tonically inhibit thalamic cells (VA) which have reciprocal connections with PRh. The connections from SNr to VA and between VA and PRh are not learned but focused (one-to-one connection pattern), meaning that the inhibition of one SNr cell leads to the thalamic stimulation of a unique cell in SNr. A dopaminergic cell (SNc) receives information about the received reward (R) and learns to associate it with striatal activities. Its signal modulates learning at the connections between cortical areas (PRh and dlPFC) and CN, between CN and SNr, as well as within SNr.

In order to test the ability of our model to perform visual WM tasks, we focused on three classical experimental paradigms (Fig. 38(B)): the delayed matching-to-sample (DMS), the delayed nonmatching-to-sample (DNMS) and the delayed pair-association (DPA) tasks. These three tasks classically consist in presenting to the subject a visual object (called the cue), followed after a certain delay by an array of objects, including a target towards which a response should be made (either a saccade or a pointing movement or a button press). In DMS, the target is the same object as the cue; in DNMS, the target is the object that is different from the cue; in DPA, the target is an object artificially but constantly associated to the cue. These three tasks are known to involve differentially IT, MTL, PFC and BG (Chang et al., 2002). In these experiments, we use four different cues (labeled A, B, C and D) and three task symbols (DMS, DNMS and DPA) that stimulate each a different cell in PRh. The corresponding cells will therefore be successively activated according to the timecourse of the trial.

Each PRh cell is stimulated by its corresponding visual object by setting the input to a value of 1.0 during the whole period. In the choice period, the input is limited to 0.5 for both cells (to mimic competition in the lower areas). To determine the response made by the system, we simply compare the activities of the two stimulated PRh cells at the end of the choice period. If the activity of the cell representing the target is

greater than for the distractor, we hypothesize that this greater activation will feed back in the ventral stream and generate an attentional effect that will guide a saccade toward the corresponding object as required in WP 3.2. We assume that such this selection is noisy, what is modeled by introducing a probabilistic rule for the delivery of reward that depends on the difference of PRh activity for the two presented stimuli.

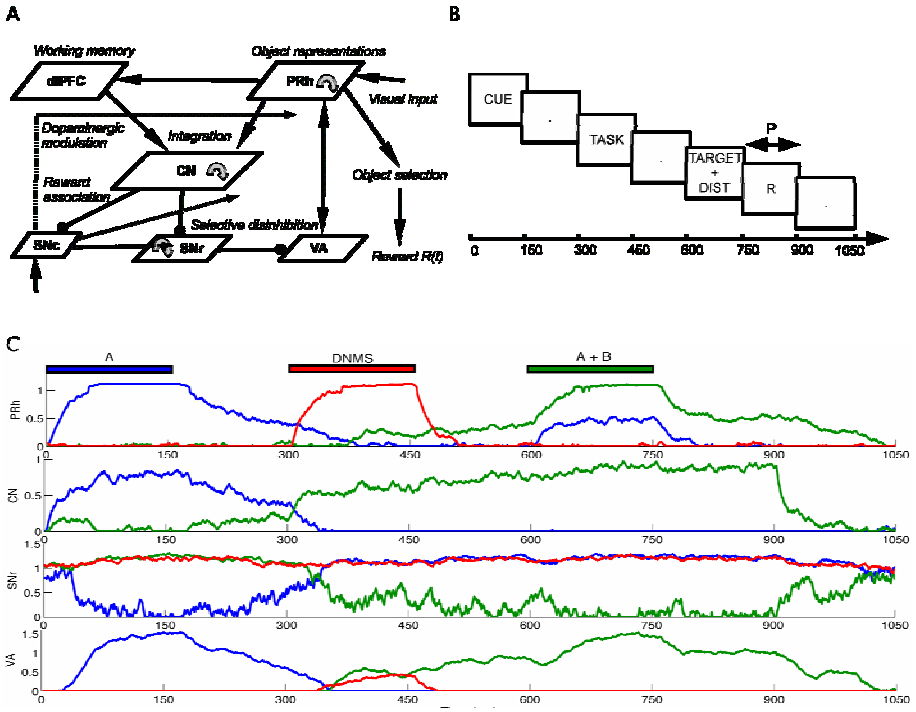


Figure 38: *(A) Architecture of the model. Pointed arrows denote excitatory connections and rounded arrows denote inhibitory ones. Circular arrows within an area represent lateral connections between the cells of this area. (B) Timecourse of the visual inputs presented to the network using DMS, DNMS or DPA tasks. (C) Temporal evolution of the activity of several cells in a network which successfully learned DMS-DNMS_AB. The activities are plotted with regard to time (in ms) during a trial consisting of A as a cue, DNMS as a task symbol and B as a target. The first row represents the activities of three cells in PRh which are respectively selective for A (blue line), DNMS (red line) and B (green line). The second row shows the activities of two cells in CN, one being selective for the pair A+DMS (blue line), the other for the pair A+DNMS (green line). The third row represents the activities of three cells in SNr which are respectively selective for A (blue line), DNMS (red line) and B (green line). The fourth row represents the activities of three cells in VA which are respectively selective for A (blue line), DNMS (red line) and B (green line).*

Fig. 38(C) shows the temporal evolution of some cells of a particular network that successfully learned DMS-DNMS_AB. The learning phase consisted of 1000 randomly interleaved trials. At the end of learning, the network was able to generate systematically correct responses which all provoked the delivery of reward. The selectivity of CN cells developed to represent the different combinations of cues and task symbols through clusters of cells. SNr cells also became selective for some of these clusters and the learned competition between them ensured that only one SNr cell can be active at the same time in this context. The temporal evolution of the activity of the cells was recorded during the course of a trial using A as a cue and DNMS as a task symbol. However, this pattern is qualitatively observed in every network that successfully learned the task and similar activation patterns occur for different tasks.
When the object A is presented as a cue in PRh (and simultaneously enters the working memory in dlPFC), it excites a cluster of cells in CN which, in this example, represents the couple A+DMS (blue line). This cluster inhibits the cell representing A in SNr which in turn stops inhibiting the corresponding cell in VA. The thalamocortical loop is then disinhibited and the two cells representing A in PRh and VA excite each other.

After 150 ms, the stimulation corresponding to the cue ends and the activity of the cells representing A slowly decreases to their baseline. At 300 ms, the object specifying the task (DNMS) stimulates a cell in PRh and enters WM in dlPFC. This information biases processing in CN so that a new cluster representing A+DNMS gets activated (green line) and disinhibits through SNr the cell in VA representing the object B, which is the target of the task. At 600 ms, when both objects A (distractor) and B (target) stimulates PRh, the perirhinal cell A only receives visual information, while the cell B receives both visual and thalamic stimulation. Consequently, its activity is higher than the cell A and will be considered as guiding a saccade toward the object B. The cell representing DNMS in SNr never gets inhibited because it has never been the target of a task during learning. The corresponding thalamic cell only shows a small increase during the presentation of the object in PRh because of the corticothalamic connection.
Three features are particularly interesting in this temporal evolution and have been observed for every network used. The first one is that the perirhinal and thalamic cells corresponding to the object B are activated in advance to the presentation of the target and the distractor. The network developed a predictive code by learning the input, context and target association. For example, the behaviour of the perirhinal cell correlates with the finding of pair-recall activities in IT and PRh during DPA tasks: some cells visually selective for the associated object have been shown to exhibit activation in advance to its presentation (Naya et al., 2003). Similarly, the behavior of the thalamic cell can be compared to the delay period activity of MD thalamic cells (part of the executive loop) during oculomotor WM tasks (Watanabe et al., 2004). The second interesting observation is the sustained activation of the perirhinal cell B after the disappearance of the target (between 750 and 900 ms on the figure) which is solely provoked by thalamic stimulation (as the WM in dlPFC still excites CN), whereas classical models of visual WM suggest that it is due a direct feedback from dlPFC (Ranganath, 2006).

The third interesting feature is the fact that the network, when only the cue was presented in PRh and dlPFC, already started to disinhibit the corresponding thalamic cell, somehow anticipating to perform the DMS task. We tested the 50 networks after learning the DMS-DNMS_AB task and presented them with either A or B for 200 ms. By subsequently recording the activity of the corresponding cells in SNr, we noticed that they all tended to perform DMS on the cue, i.e. disinhibiting the corresponding thalamic cell. This can be explained by the fact that the representation of the cue in PRh is also the correct answer to the task when DMS is required, and the projection from PRh to CN therefore favors the selection of the striatal cluster representing A+DMS compared to A+DNMS. This can be interpreted as the fact that the "normal" role of the visual loop is to maintain the visually presented objects, but that this behavior can be modified by additional prefrontal biasing (here the entry of DNMS into WM and its influence on striatal activation), as suggested by Miller & Cohen (2001).

In the remaining funding period of Eyeshots (please refer to D3.3b which is due at month 36) we envisage to further improve this model and simulate it in the context of a particular Eyeshots scenario, such as the 3D scene used earlier where a decision has to be made which of the objects is relevant for being fixated or grasped (i.e. leading to reward). In such a scenario the BG ought to learn the appropriate top-down signal.

Deviations from the project workprogramme

No deviations have been taken

WP4: Sensorimotor integration

Leader: Angel del Pobil (UJI)
Contributors and planned/actual effort (PMs) per participant: [planned]
UJI (13.4/13.4), UNIBO (10/10), UG (3/3.19), WWU (0/3)
Planned/actual Starting date: Month 1/1

Workpackage objectives

During the second year, research in WP4 has been mostly focused on Task4.2, which began in year one. The goal of such task is to generate a sensorimotor representation of objects in the peripersonal space in a dynamical way, through the practical interaction of an artificial agent with the environment, and using both visual input and proprioceptive data concerning eye and arm movements.
The results of WP4 in year one were the model framework developed for Task 4.1 and part of a data analysis aimed at modeling purposes developed at the start of Task 4.2. For the end of year two, the computational model of Task 4.2, based also on neuroscience data from WP5, was expected to be in an advanced stage of development, in order to apply it to the UJI humanoid robotic setup before the end of the task at month 30.
The modeled robot first, and the real setup later, are expected to naturally achieve very good open-loop gazing and reaching capabilities towards given target positions, and reproduce/emulate some neuroscience findings.

Task 4.3 was planned to begin during year two, for achieving at the end of the period a work-in-progress stage in which the working plan for the last year project was clearly defined. Finally, the agent will simultaneously learn to reach towards different visual targets, achieve binding capabilities through active exploration and build an egocentric, 3D "visuomotor map" of the environment. This objective requires the introduction of a sensorimotor memory module and the integration with ventral information on object identity.

Progress towards objectives

Task 4.2: Generating visuo-motor descriptors of reachable objects

Use of the model from Task 4.1

The model framework developed for Task 4.1 is the starting point for the development of Tasks 4.2 and 4.3. Its basic components and principles constitute the main building blocks of WP4, although the model had to be adapted to be suited to the requirements of the advanced behavior planned for the subsequent tasks, and the later application on a real robotic setup.
The most relevant modeling aspects that have been taken into account during year two of the project refer to the integration of visual, oculomotor and arm-motor information into a common action-oriented framework, aimed at endowing an artificial agent with a natural ability in interacting with features of its surrounding environment. Issues and concepts related to the job of the dorso-medial cortical visual stream, especially regarding its visuomotor properties, were the starting point of the computational implementation in Task 4.2, as described below. For what concerns the job of the ventral stream and its involvement in the exploration of the peripersonal space, this will be included in the description of the advancement of Task 4.3.

Resume of data analysis and computational insights

Considering the functional role of the dorso-medial stream, information regarding eye position and gaze direction is very likely employed by area V6A in order to estimate the position of surrounding objects and guide reaching movements toward them. During the second year we have **continued and improved the computationally-oriented data analysis on different V6A experiments in collaboration with the UNIBO**

partner (Chinellato et al. 2009a, Bosco et al. 2009, see also Deliverable 4.2a and Task 5.1). The new insights and UNIBO previous findings clearly indicate a critical role for 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 planning and monitoring of reaching movements, indicating that this area is in charge of executing 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 Chinellato et al. 2009a, we performed a Principal Component Analysis on the activation patterns of V6A neurons showing that a neural population of V6A can be properly modeled by a basis function approach. With our model we try to verify what computational advantages could be given by a responsiveness pattern such as that of V6A, including neurons having only visual response, neurons apparently involved mainly in motor actions and mixed neurons, activated in all phases of sensorimotor processes. Still unpublished results on the vergence/version modulation of V6A neurons have also been taken as inspiration for modeling.

Modeling of dorsal stream mechanisms for achieving visual/oculomotor/arm-motor coordination.

From the beginning of the project, basis function networks (Pouget & Sejnowski, 1997, Salinas & Thier 2000) have been considered the most important building block for our modeling approach, as explained in more detail in D4.2a. The use of basis function networks is a novelty for research works in which an artificial agent learns visuomotor coordination by interacting with its environment. Considering also different approaches such as Self Organizing Maps, we are the first to exploit real stereo vision in eye-arm coordination, realizing also a joint control of vergence and version movements. Moreover, our sensorimotor transformations are bidirectional, so that our system is the first one that learns to gaze towards its hand but also to reach where it is looking at.
The process of localization of a potential target 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. The building blocks of our framework are shown in Fig. 39.

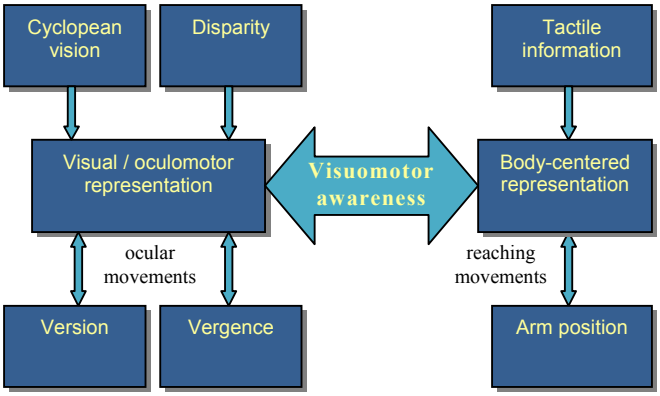


Figure 39: *Building blocks of the global 3D space representation. The body-centered 3D map constitutes the integrated visuomotor representation of the peripersonal space.*

Several possible alternatives for representing the above information can be employed. Among the possible alternatives for representing binocular information we favor the composition of a cyclopean image representation with a disparity map (under the assumption that the correspondence problem is already solved), over the option of having separate left and right retinotopic maps. Similarly, considering that we are modeling extrastriate and associative visual areas, it is plausible to assume that gazing direction is represented by version and vergence angles instead of the two explicit eye positions. 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. In coordination with WP1, we are now working on a transformation scheme that employs a log-polar visual

representation, for allowing increased visual precision in the fovea. At the other hand of the transformation scheme, we find the body-centered map of the peripersonal space required to code arm movements, which is also head-centered in our case, being the neck fixed. The map is thus accessed and updated when required, as described below.

Model details

The use of robot hardware constitutes a possible complication in the realization of a model of cortical mechanisms, and some issues that would easily be solved in simulated environments have to be dealt with more accurately considering the real world implementation. The robot is initially endowed with an innate knowledge of how to move in its environment, which is later developed and customized through exploration and interaction with visual and tactile stimuli. Following this idea, all transformations are first implemented on a computational model, which final configuration is used as a bootstrap condition for the actual experimental learning process by the robot.

Although in principle one representation should be enough for all the required transformations, the number of neurons necessary 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 look thus especially critical for performing sensorimotor transformations. 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 global structure of the model follows these principles, and is thus modular, at least separated in the visual and oculomotor component and the arm sensorimotor component (left and right sides of Fig. 39). This has also the advantage of allowing us to check for the achievement of the different capabilities separately, and accelerate the learning process.

Learning the transformation from binocular visual data to eye position consists in identifying visual targets and foveating them with both eyes, in order to associate appropriate version and vergence movements to retinal locations. As anticipated, the visual input regarding a potential target is expressed with its location in a cyclopean visual field accompanied by information on binocular disparity; we obtain in output the correspondent vergence/version movement required to foveate on the target. The transformation has been implemented with a RBF network, using a fixed retinotopic distribution of the basis function neurons. The inverse transformation from vergence/version to visual data is also possible, and its meaning is that of predicting the values of cyclopean location and disparity of the foveated point after the movement. Any difference would signal either a wrong movement execution or a displacement of the target in view.

More complex and interesting is the second learning phase, in which arm movements are introduced, as exemplified in Fig. 40. This phase is further subdivided in two stages, respectively free and goal-based. The free exploration consists of random arm movements and subsequent saccades toward the final hand position, which allows the agent to learn the transformation from joint space to oculomotor space. In the goal-oriented exploration a target object in space has to be foveated and reached. During this process, the inverse transformation having arm joints in output is learnt.

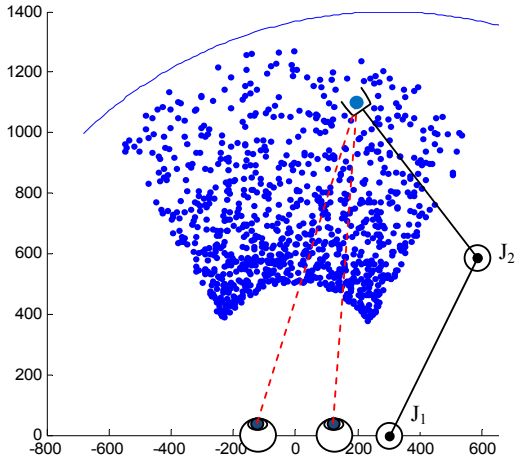


Figure 40: *Robot gazing and reaching.*

A very important issue at this stage of the process is how to define the basis function neurons. Neuroscience findings regarding area V6A are especially relevant at this stage. We mentioned before that a basis function approach is suitable to model a population of V6A neurons, and that this area includes neurons having only visual response, neurons apparently involved mainly in motor actions and mixed neurons, activated in all phases of sensorimotor processes. With our model we wanted to check what computational advantages could be given by such responsiveness pattern. For simplicity at this stage, only two arm joints were used, and no tilt movements of the eyes, so that the accessible environment is a 2D space placed horizontally in front of the subject, as in Fig. 40. This is anyway consistent with most of the monkey experiments in which activity in V6A was registered (see Task 5.2).
We simulated the different types of neurons with populations of radial basis function neurons uniformly distributed in the vergence/version space and in arm joint space, as in Fig. 41. The sort of transformations the nets learn at this stage are of the type of Fig. 42, in which the two joints are plotted as a function of version and vergence.

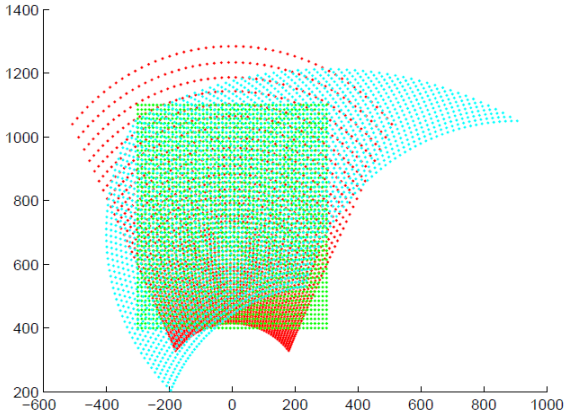


Figure 3: *Mapping of the space according to uniform distributions in a vergence/version oculomotor space (red), in a J1/J2 joint space (cyan) and in a standard Cartesian space (green).*

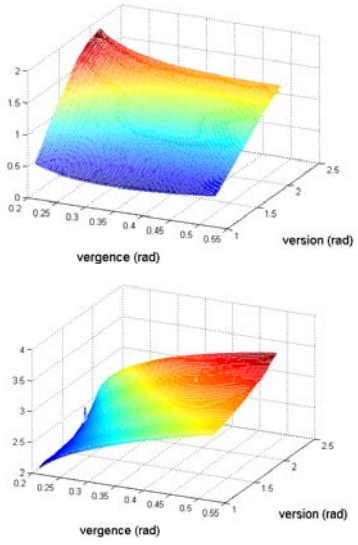


Figure 4: *Transformation from oculomotor to arm-motor space.*

Simulated results

As explained above, at this stage of the model development we wanted to achieve good performances in the learning of the transformations between oculomotor and arm motor space. This has to be done respecting, and trying to emulate, the responsiveness pattern observed in area V6A. V6A and nearby areas perform all the transformations required for a correct gazing and reaching, and for this reason, an important requirement is that the same pool of artificial neurons, centers of the radial basis functions, have to be used in the direct and inverse transformations. Natural distributions are those that follow a vergence/version space and a joint space. In order to check their suitability to model the transformations performed by V6A neurons, we trained RBF networks having the centers distributed as in Fig. 41, red and cyan graphs, for vergence/version and joint space respectively. To avoid biasing toward one or the other distribution, training and test sets were taken randomly from a Cartesian space. As depicted in Fig. 41, the ranges were taken so that the superposition between the center distributions and the training and test sets were equivalent between the oculomotor and the joint space. A further complication in the comparison between distributions is that different neuron placements and different transformations are optimized with different number of neurons and amplitudes. Hence, due to the difficulty of finding neutral parameters, equally suitable to the different testing conditions, the results, exemplified in Table 5, are expressed as qualitative assessment of performance, averaged across conditions, and not as numerical values, which justification would require a more extensive testing across parameters. To give a idea of the performance of the network, four star means that the average distance error between goal and actual position is less than 1mm, whereas one star means that it is practically always above 10cm. In any case, the trend is quite clear and rather consistent. The vergence/version distribution of neurons is reasonably good in both transformations, from oculomotor to joint space and inverse, whilst the joint space distribution is good only for the joint to oculomotor transformation. A mixed distribution, with both types of neurons, allows to obtain the best results in both transformations, even when the total number of neurons is less than in the above cases (an extensively tested configuration had 64 neurons in the V-V and J-J spaces, and 50 neurons (25+25) in the mixed space, represented in Fig. 43. A complementary test was done using forward selection from pool of V-V, J-J and mixed neurons. The results were very similar to those in Table 5, showing even bigger advantages of using mixed neural populations.

Table 5. *Performance of RBF networks with neurons distributed according a vergence/version space (V-V), joint space (J-J) and a mixed space, including both types of neurons.*

Neurons	**V-V → J-J**	**J-J → V-V**
V-V	***	**
J-J	*	***
V-V + J-J	****	****

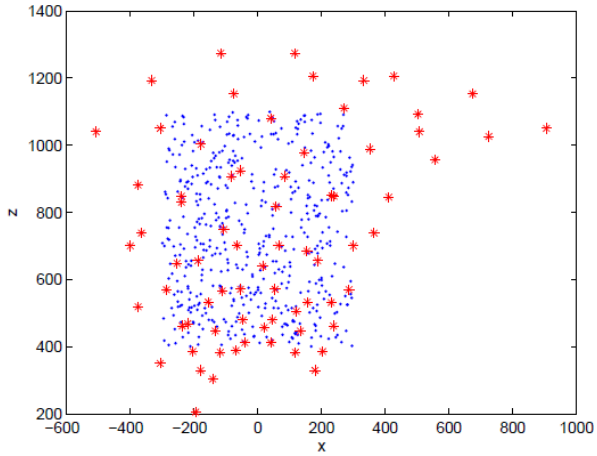


Figure 43: *Mixed distributions of radial basis functions (red stars) and test set (blue dots).*

Recent experiments [Task 5.2] show that the receptive fields of many V6A neurons seem to be indeed distributed according a vergence/version criterion. Less clear is the effect of joint space, also because of our over-simplification of the arm joint space. In any case, **our simulation supports the hypothesis that a mixed population of neurons such as that observed in V6A is especially suitable to a cortical area which contextually codes for different reference frames.** From a robotic point of view, **through the use of basis function neurons whose configuration was set according to what suggested by neuroscience data, we were able to learn very accurately the transformations between oculomotor and joint space, in a way suitable to their application to the robotic setup.**

Reproduction of psychophysical findings

The model and its robotic implementation are expected to be able to reproduce some of the experimental protocols and related effects used in WP5. Regarding psychophysical effects, we are beginning to check the model behavior in the case of the deceptive visual feedback, such as in typical experiments of saccadic adaptation (See Task 5.3). This is done by eliciting a saccade (based on vergence/version eye movement control) toward a given visual target, and providing a fictitious error on the reached final position. For the computational model, this is achieved by adding an offset to the output. On the robot, the same effect will be obtained moving the visual target as for human subjects. Analysis of how (as in the saccadic adaptation protocol) such artificial displacement of the target affects the artificial agent oculomotor and arm motor abilities can serve as a validation of the underlying model, and may help in advance hypothesis on saccadic adaptation mechanisms in humans and monkeys. So far, we were able to verify that **our model does exhibit saccadic adaptation**, altering its ability to perform correct saccades according to the deceptive feedback. The analysis of error distributions around the target point and of error vectors is also providing interesting information that we are studying together with the partner WWU. We expect to achieve even more revealing insights from applying the same protocol to the robotic setup.

Hardware setup and setting of robotic experiments.

One of the principal goals of Eyeshots 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 implementation on an actual, physical sensorimotor setup is a potential source of additional insights for the computational model, hardly achievable with simulated data.

During year two of the project, **we acquired a pan-tilt-vergence stereo head, configured it and set it up on a new torso body, together with two multi-joint arms,** endowed with a parallel-jaw gripper and a three finger Barrett Hand respectively (see Fig. 44). The first efforts of the Intelligence Robotic Lab were devoted to integrating it and interfacing it, both in hardware and software, to the previously available equipment. Among other utilities, we developed a library, which can be accessed by software or as a user interface, for the coordinated control of head movements according to a vergence/version control, as in Eyeshots objectives.

We are now working toward the application of the model on the humanoid robot, at first placing the workspace at eye level, so that only 2D eye and arm movements are required. After the 2D transformation have been successfully applied to the robot according to the model described above, we plan to extend it to the 3D space, introducing tilt movements or the head and at least one more joint for the arm. Preliminary studies with three-input RBF transformations were successful in this regard.

We describe next the steps required to implement on the robot the exploration behavior aimed at learning an implicit sensorimotor map of the environment. This is not done from scratch, as learning is bootstrapped with the weights learned during the training of the modeled network. The learning is now incremental, depending on the outcome of each action. Possible misalignments are made of two different error components, one due to the visual-oculomotor transformation and the other to the arm-oculomotor. The two error components can be estimated measuring the visual distance between the effector and the final gazing point. The use of tactile feedback upon object touching can finally constitute a master signal that allows to infer the exact magnitude of both errors. This is indeed the normal behavior of the agent, which simply always continues learning in each gazing or reaching movement towards nearby goals.

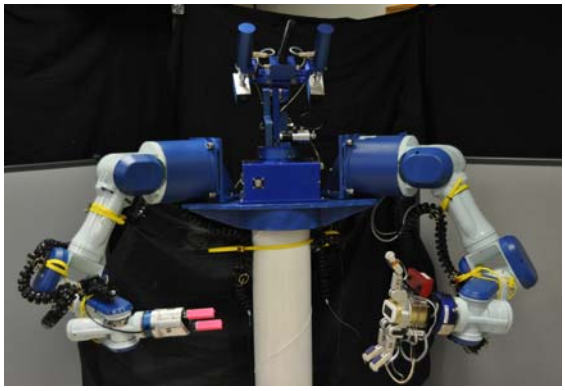
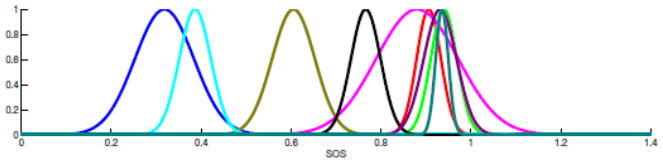


Figure 44: *Humanoid robot with pan/tilt/vergence head and arm with hand.*

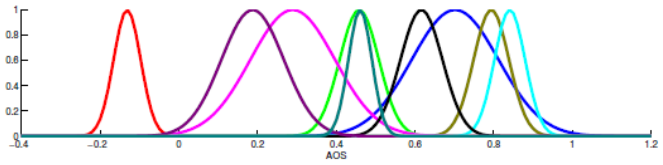
Task 4.3: ***Constructing a global awareness of the peripersonal space.***

Ventral stream: object recognition

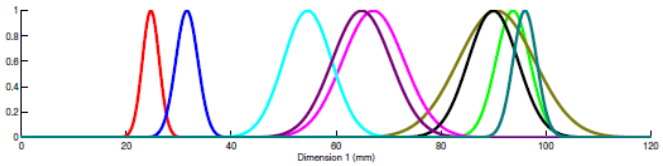
In addition to the skills of Task 4.2 in building a sensorimotor map of visual and motor targets in the nearby space, the construction of an integrated knowledge of the environment requires also the identification of objects or targets and the use of a memory of previously observed/reached objects.
An important open issue that needs to be solved is how to associate object identity to motor memory of locations. So far, we have dealt with the object recognition task to identify simple shapes according to various available features. Among others, we include visual features regarding object proportion, a sort of information that is very likely forwarded to the ventral stream by dorsal areas. Such features are the activation of Surface Orientation Selective (SOS) and Axis Orientation Selective (AOS) Neurons of the posterior intraparietal sulcus area CIP (Chinellato & del Pobil, 2008). In addition, we include the estimated size of the object three main axes, its general shape (approximated to box, cylinder or sphere) and its color, for a total of seven features. The color feature is dominant, and no other features would be required if objects had different colors. In any case, we are interested in having a robust visual object identifier, able in principle to recognize objects of the same material and thus with very similar colors.
Our tests so far have been reduced to a box universe. At first, the system is provided with a number of labeled items, and uses visual perception to associate detected features to object identity. Then, upon presentation of a novel input, it tries to recognize the visualized object, through a probability density estimator. The system decides that a target object has been recognized when there are at least five features that match one of the known classes. A feature matches if the sample is in the range $[\mu - n\sigma, \mu + n\sigma]$, where parameter n defines the tolerance of the classifier, and has been set to 2, corresponding in theory to about a 95% probability of correct classification (n = 3 would correspond to approximately 99%). If a sample is classified, it is directly added to the classifier memory to be used in subsequent tests, otherwise it is ignored. If a human supervisor is available, he is asked to label the sample so it can be added to the memory. The **average identification performance obtained so far on box-like objects of the same color is of 83%**
As an example, Fig. 45 shows object identity plotted as a function of SOS and AOS activations, and of estimated height. While for some objects these values are very informative and nearly enough to recognition, it is apparent that other objects cannot be resolved from each other only using one or two visual features. If we decide to include grasping in our experiments for Eyeshots, object weight, estimated upon lifting thanks to haptic perception of the grasped object, can also be used to distinguish and recognize objects.



(a) Results of SOS activation.



(b) Results of AOS activation.



(c) Results of height estimation.

Figure 45: *Discrimination of object identity with SOS and AOS data and estimated object height, for nine objects.*

Human-robot interaction setup.

For what concerns joint human-robot experiments, we have been exploring with the WWU partner the possibility of interfacing an eye-tracker device with the robot, thus transmitting to the system very precise information on where the human experimenter is looking at. Experimental protocols in which the robot is required to interact with the human such as for human-human experiments can be performed in this way. The experiments we plan to execute on this setup includes easy ones from the reaching and /or gazing of objects the experimenter is fixing upon or pointing toward, to more demanding tasks in which the system has to be able to perform a sequence of movements repeating with a certain delay what the human partner has done. We plan to try also experiments of "safety interaction" in which the robot is required to move away from an object that the experimenter may want to reach. Experiments on human-human interaction performed by WWU will be taken as inspiration for this final experimental step. So far **we have been working on establishing a communication protocol between human and robot according to which human intention is easy to recognize for the artificial agent, and we have been working on the software for interfacing the eye tracker with the robotic system.** Contingency plans include the possibility of using off-line data from the eye-tracker if finally it is not possible to actually integrate the two systems physically.

Deviations from the project workprogramme

None.

WP5: Human behaviour and neural correlates of neural multisensory 3D representation

Leader: Patrizia Fattori (UNIBO)
Contributors and planned/actual effort (PMs) per participant:
UG (0.5/1), WWU (16/13), UNIBO (48/48), UJI (1/1)
Planned/actual Starting date: Month 1/1

Workpackage objectives

This Workpackage is devoted to the definition and the execution of specifically-designed neurophysiological and psychophysical experiments to study the human behavior of active perception and to find neural correlates of multisensory 3-D representation. Specific results of the different WP5 tasks will be used to implement computational models developed in other WPs, providing architectural guidelines for the organization of perceptual interactions and the production of artificial intelligent systems able to explore and interact with the 3D world.
In the first year the foreseen activity was the preparation of the experimental set-up and the realization of monkey training to perform fix-in-depth and reach-in-depth in controlled conditions.

This Workpackage is devoted to the definition and the execution of specifically-designed neurophysiological and psychophysical experiments to study the human behavior of active perception and to find neural correlates of multisensory 3-D representation. Specific results of the different WP5 tasks will be used to implement computational models developed in other WPs, providing architectural guidelines for the organization of perceptual interactions and the production of artificial intelligent systems able to explore and interact with the 3D world.

Progress towards objectives

Neurophysiological experiments:
The medial parieto-occipital cortex of primates is considered a crucial area for the sensorimotor transformations that underlie the control of visually-guided arm movements (Galletti et al., 2003). Cells modulated by complex visual stimuli (Galletti et al., 1999), and reach-related cells (Galletti et al., 1997; Fattori et al., 2001), also spatially oriented (Fattori et al., 2005) are present in this brain region. Gaze position effects in this part of the brain have been found both on spontaneous activity (gaze fixation in darkness) and on visual responses (visual stimulation of the receptive field) (Galletti et al., 1995). Oculomotor signals were demonstrated during a saccade task (Kutz et al., 2003). In another study the postsaccadic activity of medial parietal neurons was found to be modulated by eye position (Nakamura et al., 1999). In all the above described tasks the fixation targets were varied on a tangent plane. We wanted to extend these studies and find what happens in the neurons of the medial parieto-occipital cortex during ocular movements and fixations in the 3D space, as well as when the animal reaches targets in the 3D space.
Psychophysical experiments:
we proceeded with the investigation of the interconnection of visual fragments and motor parameter adjustment and their parameters. We examined the influence of motor and visual parameters on object localisation obtained from saccade adaptation data. Our hypothesis that saccade adaptation modifies perceived location of saccade goals was confirmed by the experiments.
We proceeded with the understanding of the sequence of allocation of attention, direction of gaze, and movement of the arm of a human cooperative partner. This research will allow the anticipation of particular actions based on the partner's behaviour. In order to achieve this objective, we collected data both in a single actor setting and on human-human interaction.

Task 5.1: ***Role of visual and oculomotor cues in the perception of 3D space.***
In the first year of EYESHOTS, the new experimental set-up has been put in use and one monkey trained to perform the behavioural tasks required. At the end of the second year, experiments have been performed in two monkeys (*Macaca fascicularis*). All protocols have been approved by the University of Bologna (N. 94/2005-C, 22/07/2005) and conform to European laws. The monkeys were trained in a fixation-in-depth task, where they performed eye movements to fixate one of the 10 targets (LEDs) of a horizontal panel (Fig.46). The panel was located 15cm from the frontoparallel level of the eyes and 10 cm below the horizontal eye level. LEDs were arranged in 2 rows, a central and a contralateral one. Central LEDs were located 15, 30, 45, 60 and 75 cm from the frontoparallel level of the eyes and lateral LEDs were 20 cm apart each from the corresponding one in the central row.

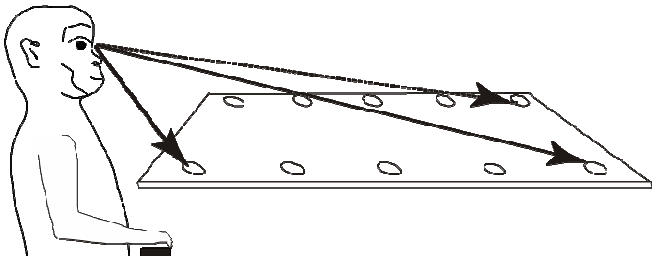


Figure 46: *horizontal panel that was used for the fixation-in -depth task to test neural activity in medial parieto-occipital cortex to oculomotor signals in 3D.*

Neural activity was recorded with a multielectrode system (Thomas Recordings) from area V6A of the medial parieto-occipital cortex and the eye position signals were acquired from both eyes with an infrared oculometer (ISCAN). At the end of this second year of the project, UNIBO has recorded a dataset of 297 V6A neurons from the two monkeys. The neuronal activity was quantified into discrete epochs of the task, related to the visual stimulation, preparation and execution of the 3D eye movement and postsaccadic fixation period. Up to now the quantitative analysis has been completed only for the first monkey. From this monkey 65 neurons were selected for analysis. Using statistical tests the effect of LED position on the activity in each of the experimental epochs was studied. We found that in about half of the neurons the activity was modulated by eye position in at least one epoch. In the great majority of these neurons the highest levels of activity were evoked for the nearest targets-LEDs. Our results so far show that neurons in V6A carry eye position/movement signals that are sufficient to form a representation of the peripersonal/reachable space (see Deliverable 5.1).
The quantitative analysis of the second animal remains to be completed. Some of the analyses are being performed in collaboration with UG. After that, we will proceed with a publication of a full paper on an international Journal with peer review with joint contributions from UNIBO and UG. Up to now, these data have been presented in an international meeting (Fattori & Galletti, 2009 [C3]).

In collaboration with UG, 2D gaze-dependent modulations in medial parieto-occipital cortex have been analysed and modelled. To study this aspect, we selected the data from about one-hundred neurons, which were tested with a fixation paradigm. These data have been submitted as joint publication with the UG unit at the IWINAC 2009 conference and lead to a joint publication (Breveglieri et al., 2009a [C10]).

In addition, in collaboration with UJI we finished the analysis the role of visual and proprioceptive guidance of reaching movements in the medial parieto-occipital cortex. Data from 150 neurons tested with a delayed arm-reaching task both in darkness and in full light have been fully analyzed. These data have been presented at international conferences during the second year of EYESHOTS.

Task 5.2: ***Link across fragments.***
This task is aimed at studying neural correlates of multisensory representation of 3D space obtained through active ocular and arm movements.
Electrophysiological recording sessions started at the beginning of the second reporting period and lasted for months, leading to the collection of neural discharges from 111 neurons from area V6A of the medial parieto-occipital cortex.
The time sequence of the reach-in depth task is sketched in Fig.47. The monkey sits in a primate chair in front of the reach-in-depth device. The monkey presses the start button placed near its belly, outside its field of view. After a delay, one of the target lights up green, and the monkey has to perform a saccadic eye movement towards the target and to adjust its vergence in order to see clearly the target light. After a variable fixation period, the fixation target turns red. This is the go signal for the monkey to release the start button and perform a reaching movement toward the fixated target. The monkey has to push the target, and to keep its hand on it until the fixation LED switches off. The monkey releases the target and performs a backward movement toward the start button to be rewarded.

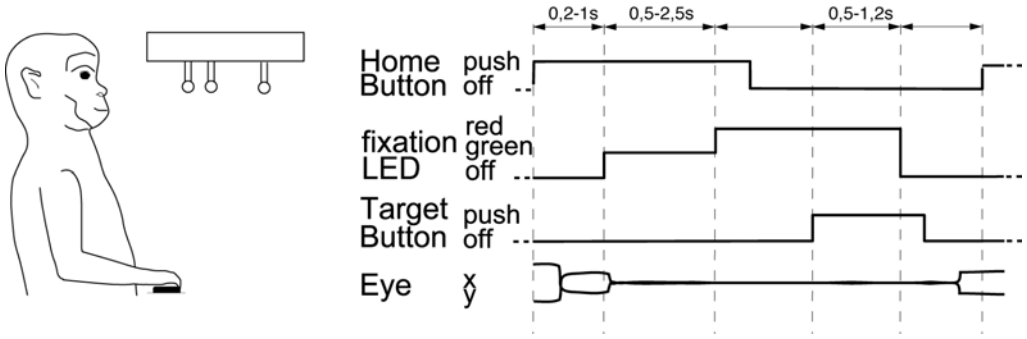


Figure 47: *timing sequence of the reach-in-depth task.*

Each studied neuron was recorded also in a fixation-in-depth task, where the monkey fixated in the same 9 positions of the panel used for the reaching-in-depth panel, without reaching any target. In the fixation trials, the same time sequence as in the reaching trials was used in the fixation trials. The change in color of the LED from green to red in fixation trials cued to the animal to release the home button and receive liquid reward without performing any reaching movement. During the tasks, the monkey was fixating the target LED from its lit up until its switching off (reaching task) or changing in color (fixation task). If fixation was broken during this interval, trials were interrupted on-line and discarded.

We analysed the discharges of neurons tested in both reaching and fixation tasks (N=111) by comparing different time epochs during the tasks. The time epochs were defined as follows: FREE: from the beginning of the trial to the light up of the LED. PERISACCADIC: from 50 ms before saccade onset to 50 ms after saccade offset. EARLY FIX: from 50 ms after saccade offset to 550 ms after saccade offset. LATE FIX: from 550 ms after saccade offset to the lit up of red LED. ALL FIX: from 50 ms after saccade offset to the light up of red LED. These task periods are common to both tasks. In the reaching task, we analysed also three execution epochs: MOV, from 200 ms before arm movement onset (home button release) to movement end (target button pressing); HOLD, from the end of forward reach (target button pressing) to 200 ms before return movement onset (target button release); RET, from 200 ms before movement onset (target button release) to movement end (home button pressing).

On this neural population, we performed a 2 ways ANOVA (factor 1 vergence, factor 2 version) and we looked for significant effects on factor 1, and/or 2 and/or their interaction ($p<0.05$).We found that a large majority of cells were modulated by ocular and/or reaching movements in 3D. Tables 1 and 2 summarizes these effects, and show how many neurons were modulatd by vergence and version respectively.

Table 6: *incidence of neural modulation in V6A in the fixation in depth task*

FIXATION In depth	**Vergence and/or Version and/or Interaction Effect**	**Only Vergence Effect**	**Only Version Effect**	**Only interaction Effect**
Perisaccadic epoch	71/111 **64%**	51/111 **46%**	47/111 **42%**	15/111 **13%**
Early Fix epoch	91/111 **82%**	65/111 **58%**	58/111 **52%**	25/111 **22%**
Late Fix epoch	89/111 **80%**	68/111 **61%**	61/111 **55%**	35/111 **31%**
All Fix epoch	96/111 **86%**	75/111 **67%**	65/111 **58%**	37/111 **33%**

Table 7: *incidence of neural modulation in V6A in the reaching in depth task*

REACHING In depth	**Vergence and/or Version and/or Interaction Effect**	**Only Vergence Effect**	**Only Version Effect**	**Only interaction Effect**
Perisaccadic epoch	67/111 **60%**	46/111 **41%**	37/111 **33%**	21/111 **19%**
Early Fix epoch	91/111 **82%**	60/111 **54%**	57/111 **51%**	27/111 **24%**
Late Fix epoch	95/111 **85%**	77/111 **69%**	53/111 **48%**	39/111 **35%**
All Fix epoch	103/111 **93%**	80/111 **72%**	66/111 **59%**	42/111 **38%**
Mov epoch	93/111 **84%**	79/111 **71%**	63/111 **57%**	42/111 **38%**
Hold epoch	92/111 **83%**	76/111 **68%**	57/111 **51%**	39/111 **35%**
Ret epoch	82/111 **74%**	61/111 **55%**	50/111 **45%**	19/111 **17%**

Figs 48-50 show exemplary cells in V6A showing vergence and/or version modulations in fixation periods and/or in arm-movement related periods.
The analysis of the population is currently going on and we are planning to present these data to international meetings.

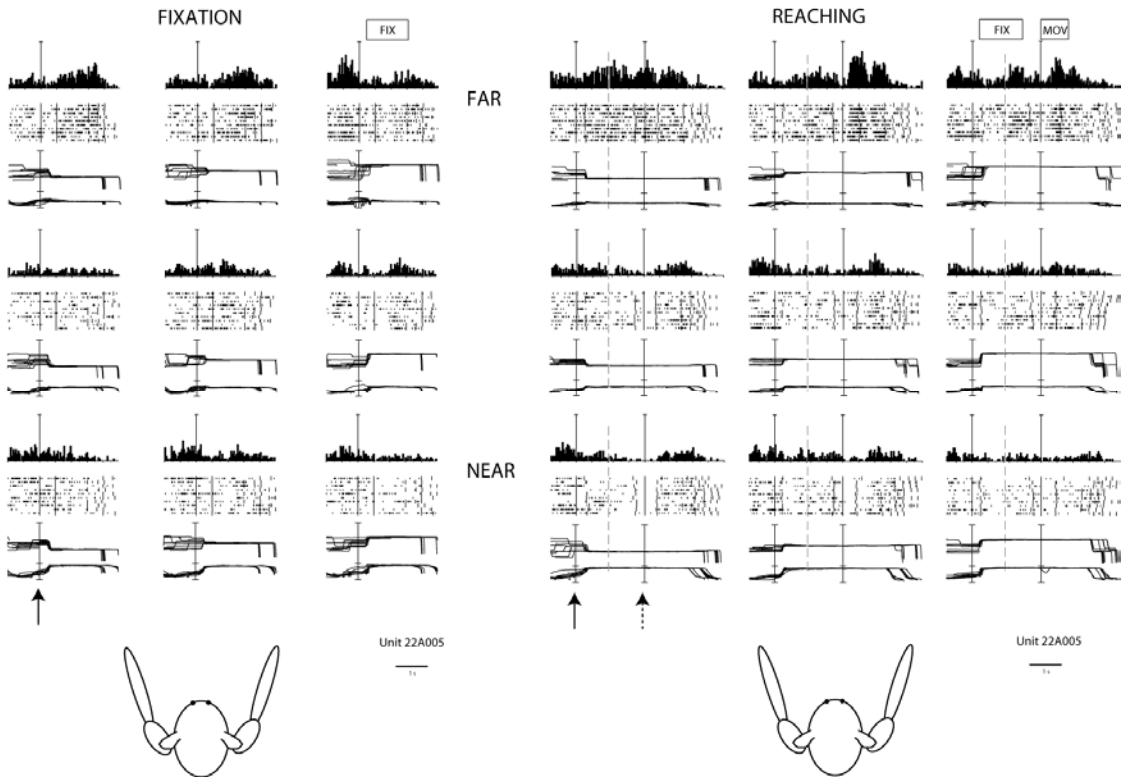


Figure 48: *example of a V6A neuron modulated during fixation epochs (both in reaching and in fixation tasks) and in the reaching epochs.*
Left: neural discharges in the fixation task; right, in the reaching task. Activity is aligned on target presentation for fixation task (arrow) and twice for the reaching task: on target presentation (continuous arrow) and on onset of the forward reaching movement (dashed arrow). Discharges are located from top to bottom, from far and near targets. Each row indicates different vergence angles (8° top, 13° center, and 18° bottom). Each column indicates different versions: left (-15°), center (0°), right (+15°).

The neuron shown in Fig. 48 is sensitive to vergence, both in the eye-related epochs (FIX) and in the arm-related epochs (MOV), showing a congruent preference for the 2 effectors landing on the farthest targets. In other cells, the neural activity was influenced by depth only in one of the 2 effectors. The example reported in Fig. 49 is a cell sensitive to vergence (with a preference for high vergence angles) and version (with a preference for right fixations) in both tasks but the reach-related activity was not present for any of the reached positions (epochs MOV and FIX).
In the cell reported in Fig. 50, the discharge is present only in HOLD period and in RET movement, and this discharge is modulated in depth, with a strong preference for farthest targets where the hand was held (HOLD) and from where it came back (RET). No discharges were evoked in the fixation tasks for all the target positions tested. Evidently, this cell is not reached by vergence/version information but only by postural (proprioceptive) or motor-related signals from the performing arm, contrary to the cell in Fig.49, where no arm-related signals seem to be processed, at least in the working space explored.

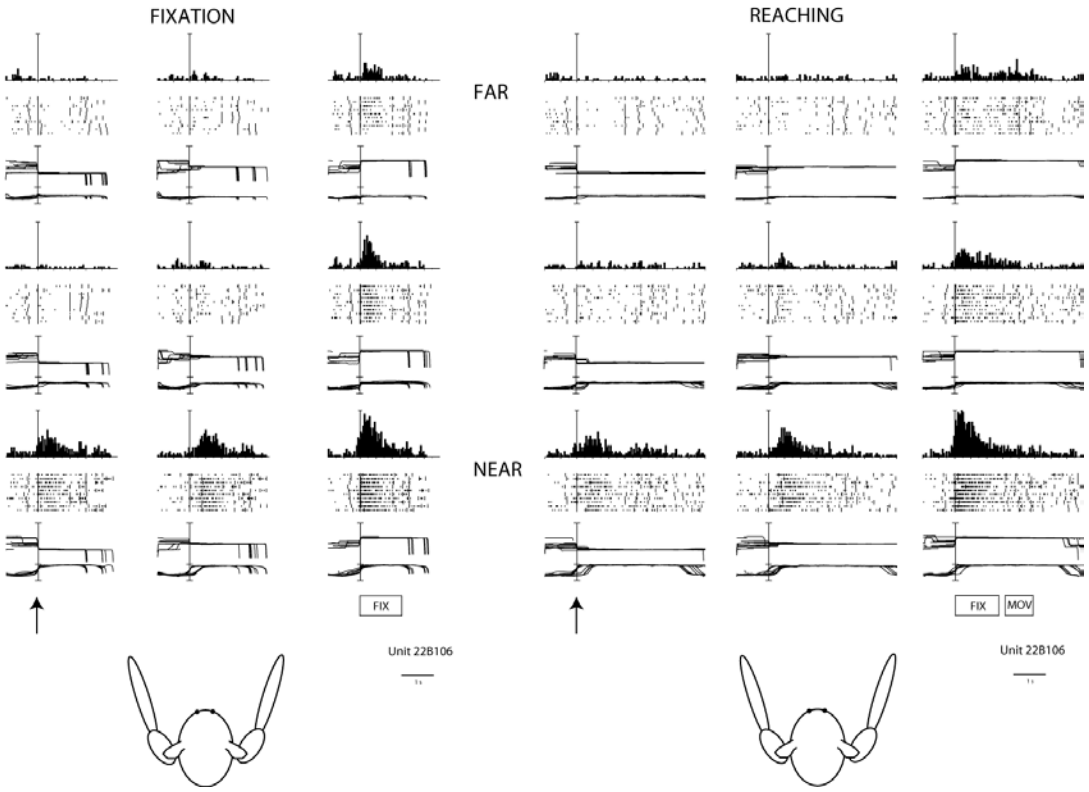


Figure 49: *example of a V6A neuron sensitive to fixations in the near space and not sensitive to the depth of the reaching movements. Activity is aligned on target presentation (arrow) for both tasks. All conventions are as in Fig. 48.*

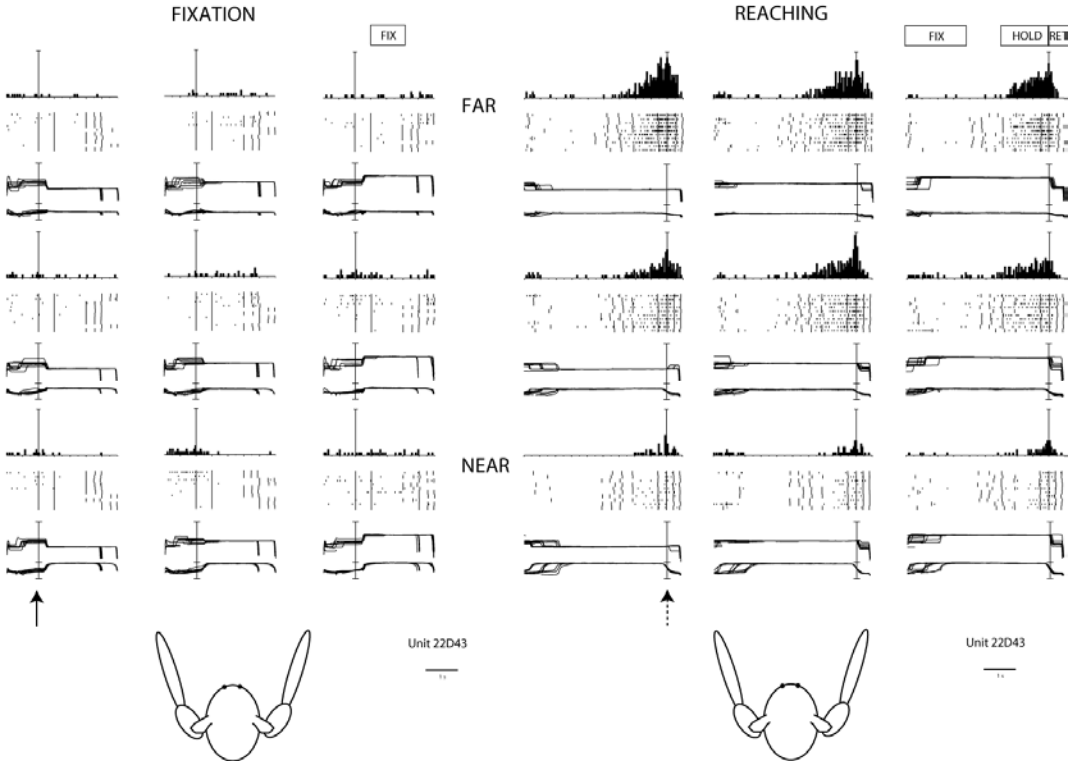


Figure 50: *neuron not sensitive to eye positions nor eye movements in depth, but sensitive to postural information and motor/related information. Activity is aligned on target presentation for fixation task (continuous arrow) and on onset of return movement for the reaching task (dashed arrow). All conventions as in Fig. 48.*

We collaborated with UJI group in the analysis of single-cell data regarding reaching experiments with different spatial/retinotopic position of the targets. The results of these analysis have been submitted as joint publication with the UJI unit at the IWINAC 2009 conference and received an award as “award José Mira to the best promising paper” submitted to IWINAC 2009 (Chinellato et al., 2009a [C8]).

Task 5.3: Motor description of fragment location

In Task 5.3 the influence of motor parameters on fragment location is examined. In the paradigm of saccadic adaptation a modification of motor parameters is evoked by the introduction of an artificial visual error after every saccade. In an experimental paradigm a subject is performing a saccade to a visual target, which is displaced during the eye movement of the subject. The retinal error experienced by the motor system hence consists of an externally controlled part and the endpoint error of the saccade. Systematically occurring errors evoke plastic changes of the saccadic gain, which leads to an effective shortening or lengthening of the saccadic amplitude. To examine the nature of this motor learning and therefore the nature of the coding of oculomotor parameters different aspects concerning saccadic adaptation are examined. The first two studies focused directly on the transfer of motor parameter changes to fragment location. In these two studies, visual fragments were located after motor parameter adaptation. Study I focused on the interplay of features of the stimuli evoking saccades and the fragments. Whereas in this study the fragments shown before a saccade were located right after the saccade, study II focussed on permanent effects of saccadic adaptation. Here fragment location was tested in fixation condition. Both studies were preliminary presented in the first period. For details also see deliverable D5.3a. Both studies are published now. **Study I is now accessible as "Mislocalization of flashed and stationary visual stimuli after adaptation of reactive and scanning saccades" in the Journal of Neuroscience Volume 35, Issue 29, study II is in press in the Journal of Vision with the title "Motor signals in visual localization".** From the knowledge that both, trans-saccadic as well as saccade independent location of fragments are affected by saccadic adaptation, three more studies on the spatiotemporal nature of saccadic adaptation were performed.

Study III deals with the time-discreteness of saccades and fragment location in contrast to the continuity of visual perception. Whereas visual perception is a continuous quale, saccades are time-discrete events, which are modified by time-discrete error signals appearing right after each saccade. Although saccades are very stereotyped, their endpoints are scattered due to noise. Therefore, every visual error is corrupted by a certain amount of insecurity. For the accurate location of fragments in visual space a strategy estimating the fractions of noise and actual visual error in the perceived retinal error is inevitable. To test the existence of such an estimator a saccadic adaptation experiment was performed, in which the retinal error was varied in mean error size as well as in consistency. If an estimator of consistency exists, inconsistent visual feedback should lead to an underestimation of the actual visual error and therefore, adaptation should be attenuated via the lowering of consistency. Figure 51 presents the main result of the study. The bars represent the gain change in deg/trial for every condition. Figure 51a shows the over all subject average for inward adaptation in the three consistency conditions. The different shades show the three consistency conditions with a variance of 0 degree, 2 degree and 4 degree standard deviations of the introduced inconsistency of the visual feedback. Decreasing consistency attenuates gain change. The analogue results for outward adaptation are depicted in Fig.51b. For outward adaptation the introduction of inconsistency leads even to an inversion of gain change, i.e. in the four degree standard deviation condition in average inward adaptation is observed. Figure 51c compares the observed effects to a baseline condition without visual feedback, where also inward gain change occurred. It is clearly visible, that inward adaptation as well as outward adaptation are attenuated via the lowering of consistency. **Study III is currently in press in the Journal of Neurophysiology with the title "The influence of the consistency of post-saccadic visual error on saccadic adaptation".**

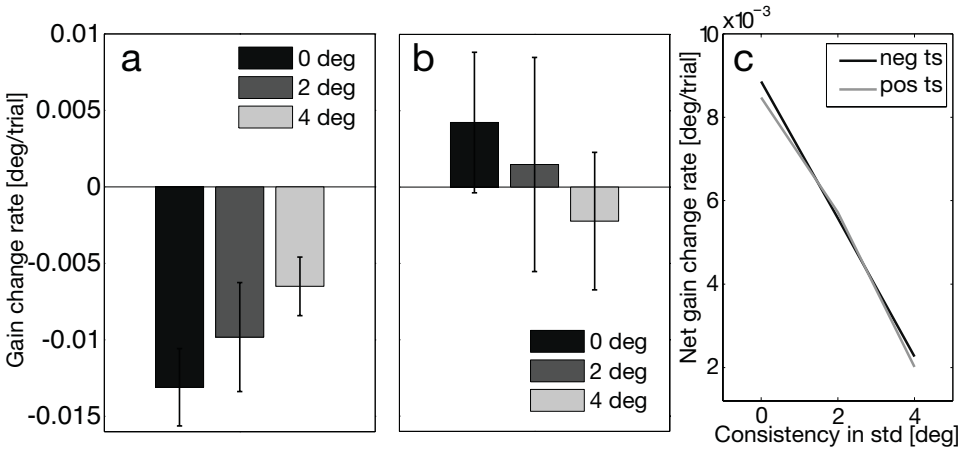


Figure 51: *a) Gain change in deg/trial averaged over all subjects for inward adaptation in the different consistency conditions. A clear attenuation of gain change with increasing inconsistency is visible. b) Gain change in deg/trial averaged over all subjects for outward adaptation in the different consistency conditions. With decreasing consistency again the gain change is attenuated, for low consistency even an inversion of gain change is visible. c) Gain changes normalized to gain change in a baseline condition without visual feedback.*

Study IV investigates the integration of the retinal input into the global extrapersonal space. The input of the retina is embedded into the three dimensional space via a series of coordinate transformations. A favoured method for the transformation from retinal information to a representation of the extrapersonal space are gain fields. The retinal input is modulated by an additional external signal, such as the eye position. The study looks for such eye position gain fields in saccadic adaptation. Whereas eye position effects were found in several visual areas no eye position effects are known in saccadic adaptation so far. After proving the close interconnection of adaptation and the localisation of fragments, the search for eye position dependencies in saccadic adaptation was a logical consequence to explore the spatial properties of saccadic adaptation. The influence of adaptation on neighboured positions was tested on several positions equally distributed on a horizontal line. A clear eye position dependency was found. In Fig.52 inward adaptation was conducted at five positions, with position three at a central position. Each of the five positions was adapted in a separate session, in which gain changes are tested for all five positions in a pre-post analysis. The bars show the mean gain change of six subjects for every test position. They are positioned analogue to the test positions. A clear eye position dependence was found, which is contrastable to gain fields. Whereas for more eccentric positions a linear decay is visible for the contralateral test positions and even a decrease evolves at more eccentric positions (Fig. 52b and 52d) for the central adaptation position the gain profile is flat. In further experiments analogue tests for outward adaptation and a vertical arrangement of test positions are preformed. Data collection for this experiment is still ongoing.

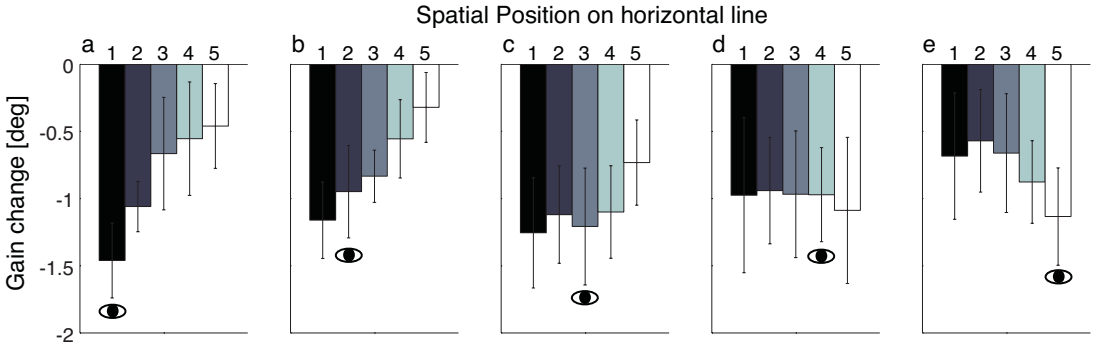


Figure 52: *Eye position dependent gain change tested on five positions on a horizontal line. a)-e) show results for adapted positions 1-5. The eye denotes the adapted position. Every sub figure shows data from all test positions. The bars are arranged analogue to the test positions. For the more eccentric adaptation positions (a), e)) a decay in gain change occurs for less eccentric and contralateral positions. For b) and d) an increase for more eccentric positions completes the linear gain profile. For the central position the gain profile is flat.*

In study V this knowledge is extended to the localisation of fragments. In congruency study I the transfer of gain change to the localisation of fragments is tested for reactive and scanning saccades. Reactive saccades are elicited by suddenly appearing targets. Scanning saccades are executed within a group of targets which are constantly visible. The test positions are arranged on a rectangular frame. The lower left corner is adapted and gain change as well as mislocalisation of fragments is tested at all four positions. The experiments are performed for reactive as well as scanning saccades, in congruency with study I. The results are depicted in Fig. 53. The black circles represent the amount of gain change at the four test positions, the grey shade shows the standard deviation. A clear decay in gain change is visible at the positions distant from the adaptation position. Data collection for this experiment is finished and a manuscript is almost ready for submission.

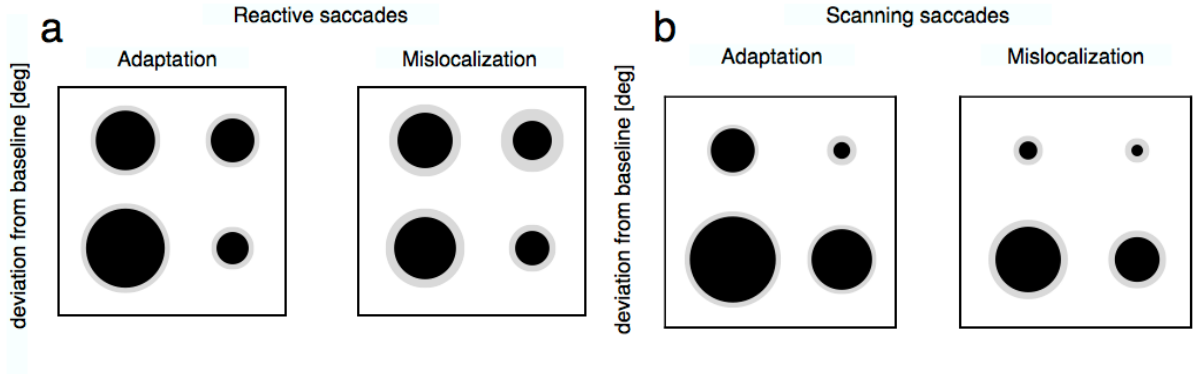


Figure 53: *a) Gain changes and mislocalisations for reactive saccades at the four test positions after adaptation of the lower left corner. A clear decay in gain change as well as localisation error is visible for more distant test positions. b) Analogue result for scanning saccades. A comparable decay in gain change as well as localisation error is visible.*

A cooperation between UNIBO and WWU has started to investigate the link between localisation and saccadic adaptation in monkeys. The experimental setup uses software from WWU for the visual stimulation and the localisation report and software that is integrated with the lab setup for eye movement control and reward in Bologna. Report of perceived fragment location by the animal will be collected with a touch screen. Scheduling of the main experiment has been delayed from the original project planning, however, because collaborative experiments between UNIBO and WWU on the influence of attention in area V6A were advanced and took precedence. These experiments were planned to document that attention modulates the firing of V6A cells and that V6A is an area that combines control processes for allocation of attention, direction of gaze, and movement of the arm.

Attention is important for providing the link across single visual fragments. Our study intended to measure the influence of covert attention toward different parts of the visual world in neurons of area V6A. With respect to the link across fragments, attention is used to select targets in a visual scene for prioritized processing and for preparing appropriately directed actions. Overt deployment of attention is seen in the directing saccadic eye movements to salient or task-relevant parts of the scene, but attention can also be deployed covertly, without any visible motor activity. Covert orientation of attention is done by internally modulating the processing of information in visual cortical maps, and by selecting parts of the scene to receive increased processing resources. The selection of the part of the scene to receive attention, i.e. the control of the focus of attention, is driven by the saliency of the stimuli and by the requirements of the task that is currently performed. It is closely related to the motor actions that are to be performed on the selected targets, in particular to the preparation of eye movements. The initiation of a saccade, for instance, is preceded by a mandatory shift of attention towards the saccade goal (Hoffman & Subramaniam, 1995; Kowler et al., 1995; Deubel & Schneider, 1996). The neuropyhsiological mechanisms of the deployment of attention are linked to the mechanisms of selecting a saccade target and preparing the saccade, even for covert attention shifts (Bisley & Goldberg, 2003; Moore & Armstrong, 2003; Cavanaugh & Wurtz, 2004; Ignashchenkova et al., 2004; Hamker, 2005; Thompson et al., 2005). In other words, this shift of attention is generated by feedback signals from cortical oculomotor areas onto perceptual areas.

However, the link between attention and goal-directed motor action is not confined to the eye movements. Also the preparation of reaching movements is paralleled by a shift of attention to the goal of the reach (Castiello, 1996; Deubel et al., 1998). It has also been demonstrated that attentional selection for simultaneous reaching and eye movements to different targets shows some degree of independence between the two, such that both goals can receive processing benefits (Jonikaitis & Deubel, 2009). Thus, one might expect that, similar to oculomotor areas that provide signals for overt and covert shifts of attention, also cortical areas that are involved in the generation of arm movements may contribute to attentional shifts. The medial parieto-occipital area V6A could be one of these areas, because V6A contains arm-reaching neurons and acts as a bridge between visual processing and arm motor coding (Galletti et al., 2003; Fattori et al., 2005; Marzocchi et al., 2008).

Though a large amount of evidence demonstrates that spatial attention modulates the neuronal response to a stimulus, much less are the evidence that spatial attention modulates the ongoing activity of a neuron. A wealth of experimental evidence demonstrates that the ongoing activity of cells in a huge number of cortical areas is modulated by the direction of gaze. Gaze modulation could actually be the effect of the overt orientation of the spotlight of attention, which moves with the gaze, but this is generally interpreted as a oculomotor effect rather than an effect of spatial attention modulation. Area V6A contains a high percentage of gaze-dependent neurons (Galletti et al., 1995; Galletti et al., 1996) and we advance the hypothesis that the gaze modulation could be an epiphenomenon of the attentional process, which in fact could be the actual factor modulating neuronal activity. To demonstrate the validity of this hypothesis we have to disengage the attention from the point of fixation, that is we have to produce a shift of the covert attention, and demonstrate that neural modulation is still present without any concurrent shift of gaze direction. Accordingly, in this work we checked whether the ongoing activity of single cells in V6A was influenced by the shifts of covert attention without any concurrent shift of the direction of gaze.

We performed extracellular recordings on single cells of area V6A of 2 Macaca Fascicularis. Animals were trained to fixate in the straight-ahead position on a light-emitting diode (LED), in darkness, while pressing a button located outside their field of view. The monkeys were trained to maintain the gaze in the straight-ahead position all throughout the trial. Their fixation was checked using an infrared oculometer. The animals

had to detect a target (5 ms- red flash) in one out of several peripheral positions and to respond to it by releasing the button. The target position was cued by a yellow flash (150 ms) presented in the same location as the target but preceding the target onset by a variable interval of 1000-1500 ms. Although we were aware of the difficulty to hold the spotlight of attention on cue location for 1-1.5 s, we chose such a long delay to have the possibility to temporarily dissociate a possible visual response to the cue from an attentional one. The cue prompted the monkeys to covertly displace attention towards the cue from an attentional one. After target detection, the monkeys had to shift the attention back towards the straight-ahead position because it had to detect a change in color of the fixation LED, and to report detection by pressing the button again. We calculated the average discharge rate of each cell during fixation before the cue onset (baseline activity) and during different time epochs after the cue onset. Out of 92 tested cells, 25 (27%) were modulated by this task (two-way ANOVA [factors: epoch x target position] and multiple comparisons with Bonferroni correction, $P<0.05$). Of these task-related cells, 56% were briefly tuned in the interval 50 - 200 ms after the cue onset (passive visual response) while higher percentage of cells were tuned during epochs starting well after the cue onset (200 - 500 ms after cue onset (70%). Moreover, 62% of task-related cells were tuned when attention was shifted back to the straight-ahead position after target detection. This modulation cannot be ascribed to any appearance of a visual stimulus, as it was continuously present and continuously fixated by the animal. Figure 54 shows the discharge of all V6A cells activated when the attention was shifted toward the periphery, after the presentation of the cue in that spatial location. For comparison, the neural discharge of the same population is also shown when the attention is shifted in the opposite direction. It is evident a sustained activity of the cells that lasts more than 500 ms, that is well beyond the cue visual stimulation. We suggest that this sustained activity is the result of the displacement of the spotlight of attention.

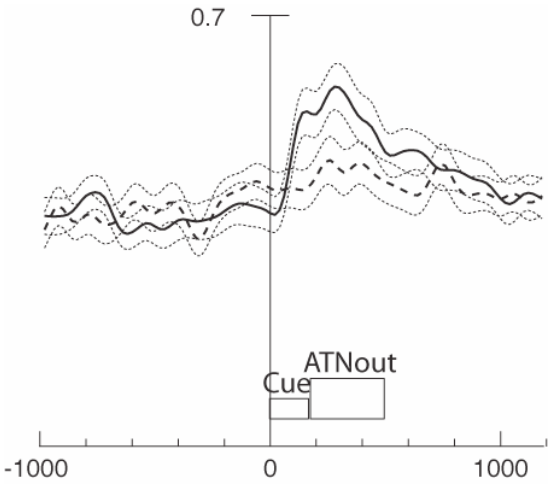


Figure 54: *population data showing V6A responses to covertly attending in the preferred (continuous line) and in the opposite (dashed line) directions.*

In other words, the medial posterior parietal area V6A contains neurons that reflect internal orientation mechanisms. These neurons can locate or select peripheral targets to be reached and grasped by the animal. These results will be available to WPs 3 and 4 as a data base for further model development and/or testing.
A first report of early data from these experiment has been presented at the 2009 Neuroscience Meeting held in Chicago (USA) (Breveglieri et al., 2009b [C13]). The data collection is now finished and the whole results are currently being evaluated and prepared for a joint publication between WWU and UNIBO.

In summary, in Task 5.3 we proceeded with the investigation of the interconnection of visual fragments and motor parameter adjustment and their parameters. We examined the influence of motor and visual parameters on object localisation obtained from saccade adaptation data. Our hypothesis that saccade adaptation modifies perceived location of saccade goals was confirmed by the experiments. Spatial and temporal properties of saccadic adaptation were explored. Detailed results were reported in deliverable D5.3a at month 15.

Task 5.4: ***Predicting behaviour and cooperation in shared workspace***
This task focuses on the understanding of the sequence of allocation of attention, direction of gaze, and movement of the arm of a human cooperative partner. This research will allow the anticipation of particular actions based on the partner's behaviour. In order to achieve this objective, we collected data both in a single actor setting and on human-human interaction.

Single actor setting experiments

Two single actor setting experiments were conducted. For both experiments we measured eye movements with the Eyelink II eye tracker system (SR Research Ltd., Mississauga, Ontario, Canada).

Study I: Preliminary data of the first experiment were presented at the end of the first reporting period. The collection and analysis of data was then completed during this second reporting period. The purpose of the experiment was to explore the hypothesis that other's gaze direction can be used to predict the behaviour of the other person. In addition, the temporal advantage of using the other's gaze direction to predict the behaviour was quantified. The experiment required the participant to watch a series of movies in which an actor performs gaze movements and reaching arm movements towards a set of targets. The participant had to predict as soon as possible the to-be-pointed-at target by looking at it. These eye movements were recorded with the eye tracker. The movies could differ in two aspects. The gazing behaviour of the actor could be visible or occluded and the targets could be as well visible or occluded. Data showed that the main factor influencing participants' responses was the availability of the gazing behaviour of the actor. The visibility of the targets had an effect on the spatial accuracy only. Actor's gaze triggered rapid and accurate responses towards the target objects, which were accurately identified when the actor's arm was still at the beginning of its trajectory toward the target object. When the actor's gaze information was not available, the participants could still predict which target object was relevant in a particular trial by relying on the actor's hand movement only. In this case, participants' gaze still leaded the hand movements of the actor, but was comparatively slower and less accurate. Figure 55 represents the timing of target identification: by using the gaze information the targets were identified with a temporal advantage of 200-250 ms with respect to the conditions in which no gaze information was available. In sum, other's gaze direction can be used advantageously as a predictive cue about the final location of a pointing movement and can be complemented by the kinematic cues provided by the hand movement.

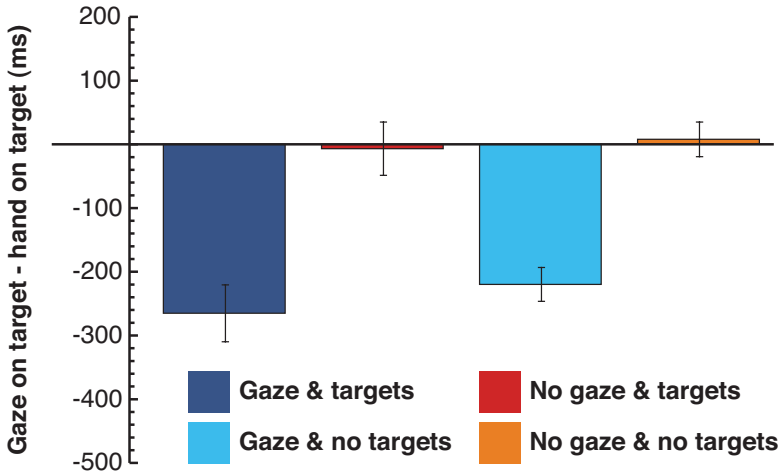


Figure 55: *Bar chart representing the timing of target identification. When the actor's gaze information was available (blue bars), the target object was identified 200-250 ms before the actor's hand reached the target. On the other hand, when gaze information was not available (red/orange bars), the target object was identified at the moment the actor's hand touched the target. Error bars denote standard errors of the mean.*

Study II: The second experiment was designed to further explore the relation between gaze behaviour and arm movements and its influence on the allocation of attention. For this purpose we ran an experiment based on a variation of the standard cueing paradigm that is generally used to study covert or overt shifts of attention. Specifically, we used a gaze cueing paradigm to study the overt orienting (involving eye movements) of attention. Typically, participants perform unwanted saccades in the direction of the actor's gaze when an instruction cue actually instructs them to saccade in the opposite direction. In the present experiment, we composed our stimuli in such way that both a gaze direction change and a small hand movement were used as distracter stimuli. Participants were asked to make a speeded saccade to left or right of fixation, as indicated by a centrally presented instruction cue. At different time intervals with respect to the instruction cue, a distracter stimulus was presented. The distracter stimuli could be of four categories. The averted eye gaze or the small hand movement were presented in isolation or in combination with each other. When presented together, their direction was either matched (both were directed towards the same target) or unmatched (gaze was directed towards one target and the hand movement towards the opposite one).

Ample evidence supports the idea that social signals, such as eye gaze, influence our voluntary eye movements. However, people move their eyes constantly and most of these eye movements are irrelevant in a context of human-human interaction. It is thus to be expected that even stronger shifts in overt attention should be induced by eye movements conveying a potentially relevant action. We hypothesised that any eye movement performed in conjunction with a hand movement should be more highly rated as a potentially relevant action than either an eye movement or a hand movement presented in isolation. If so, the distracter conditions irrespective of the direction of the small hand movement should show a larger effect on the amount of unwanted saccades in the direction of the actor's gaze.

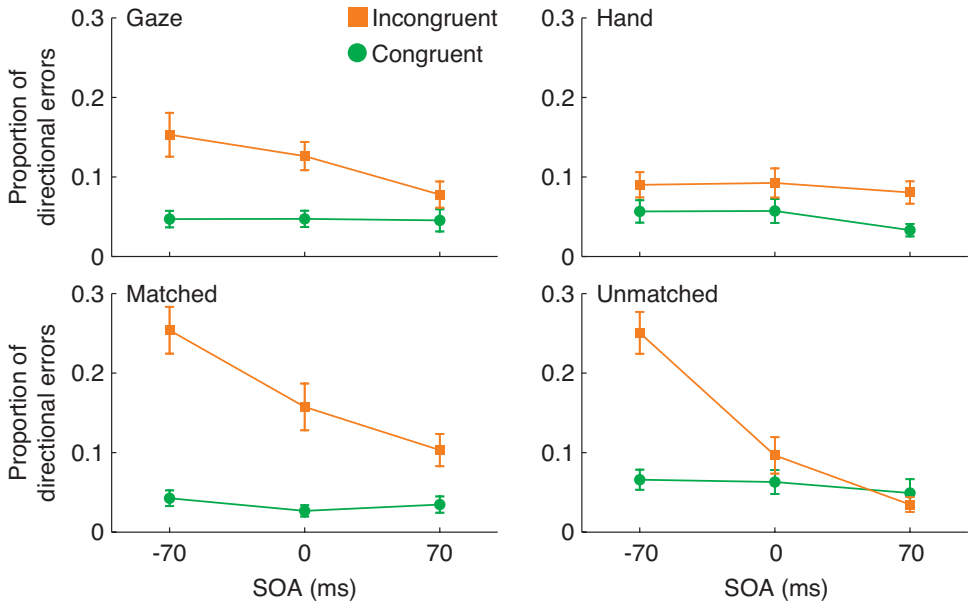


Figure 56: *Mean proportion of directional errors as a function of the stimulus onset asynchrony (SOA) made for the congruent and incongruent trials in the four distracter conditions. Error bars denote standard errors of the mean.*

Our data show that, as previously reported, gaze and hand cueing were effective at triggering the saccades in the opposite to the intended direction (directional errors). A stronger gaze cueing effect was, however, observed when the gaze and hand cue were presented in conjunction and the proportion of saccades following the gaze cue increased irrespective of the small hand movement direction. Figure 56 represents the proportion of directional errors in all conditions. The congruency effect is defined as the difference between the incongruent and the congruent conditions. It is evident that this effect is largest in the matched and unmatched conditions. We conclude that the mere presence of a sudden hand movement might have been interpreted as a sufficient indication of a forthcoming relevant action that consequently enhanced the

saliency of the directional cue provided by the gaze. These findings thus suggest a process that prioritises potentially relevant actions to which the visual system automatically responds. In addition, this finding can be implemented as one of the heuristics for an optimal human-robot interaction. The robot should not reflexively orient to every eye movement of the human cooperation partner, but only to those eye movements that lead to a potential action. Our study suggests that eye movements performed in conjunction with a hand movement could be thus classified as potentially relevant actions to which the robot should respond.

Human-human interaction

The human-human interaction experiment was conducted on the setup developed during the first reviewing period that is based on two ViewPoint eye tracker systems (Arrington Research Inc., Scottsdale, AZ). This setup allows the simultaneous recording of eye movements of two interacting participants.
Looking in the right place at the right moment is particularly important while executing movements in coordination with another person. Being able to predict outcomes of other's movements is even more crucial. We addressed this issue by exploring the gaze behaviour of pairs of participants involved in a simple cooperative task. The two participants were facing each other and each of them had to move an object in the vertical plane around an obstacle and make contact with the object of the other participant. In Figure 57, four video-frames taken from the recording of one participant show the execution of the movement. Eye movements were simultaneously measured in both participants.
A stereotypical gaze behaviour was observed: (1) at the start of each trial a fixation was directed towards the own object (see Fig. 57, top left); (2) fixation was kept on a central location of the setup (see Fig. 57, top right); (3) saccades were then regularly directed towards the partner's object in the terminal phase of the movement prior to the contact between objects (see Fig. 57, bottom left); (4) the gaze followed the object until contact was made (see Fig. 57, bottom right). These saccades started approximately 300-400 ms before contact. In addition, we measured a condition in which one of the two participants had the freedom to determine the contact location between objects and the other participant had to comply with this behaviour. In this condition we observed an earlier initiation time of the object oriented saccade. The timing of object oriented saccades was modulated by the predictability of the contact location.

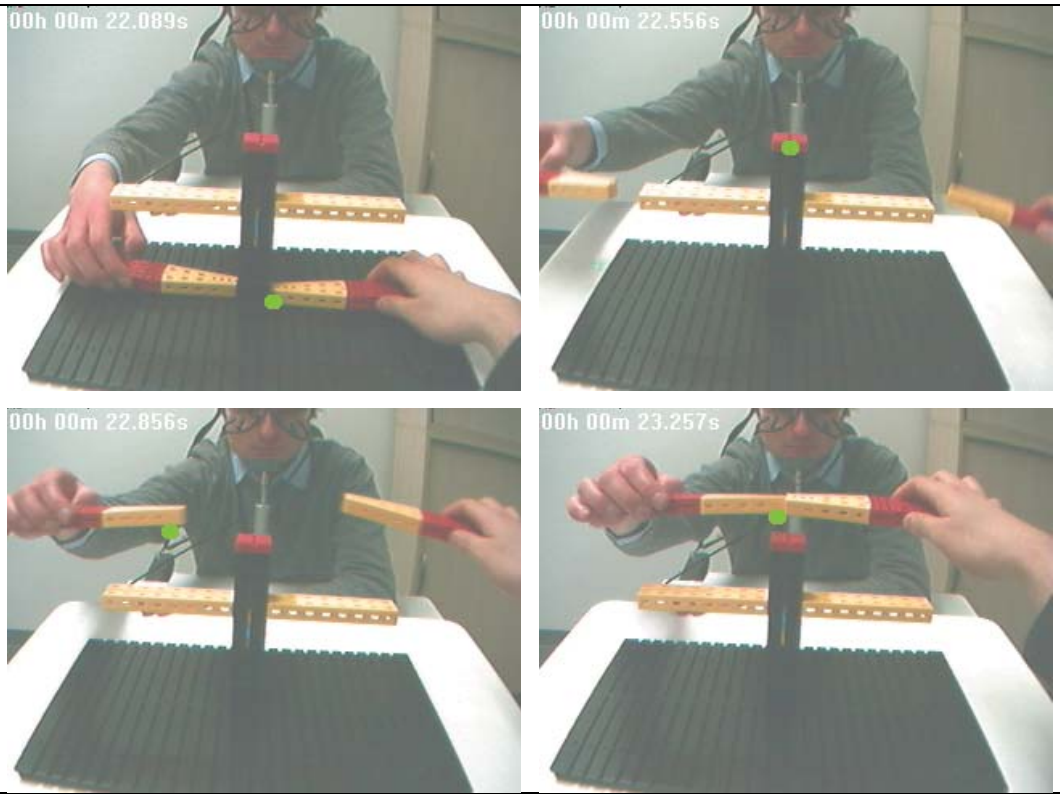


Figure 57: *Four video-frames showing the execution of the requested movement: the two participants had to move an object in the vertical plane around an obstacle and make contact with the object of the other participant. This video-frames were extracted from the recording of one of the participants. The green dot in each video-frame indicates the gazed location.*

Our data suggest that the stereotypical gaze behaviour seems necessary to establish a closed loop between the two participants that allows a coordinated fine-tuning of the joint interaction. When both participants jointly adapt their behaviour for the achievement of the final goal, fewer resources are needed for a successful interaction. The expectations that a human actor has about the cooperation partner influence the deployment of attentional resources.

The data of the experiments in a single actor setting were needed for milestone M9.ante at month 18. The milestone was reached as expected. Our hypothesis that gaze tracking can be used to predict allocation of attention and behaviour prediction was confirmed by the experiments. The data of the human-human interaction experiment will be needed for the achievement of the milestone M9 at month 27. Detailed results will be reported in deliverable D5.4 (preliminary report) at month 27.

Deviations from the project workprogramme

No major deviations from the planned experimental work have been done. Some steering of the activities has been decided, in order to follow the suggestions of the reviewers during the first reviewing meeting. Thus, in accordance with the coordinator, we decided to adapt the planned experiments in order to collect experimental evidences in the direction of bilateral interactions between perceptual and motor processes. In this line, rather than proceeding with passive visual stimulations, we opted for a more dynamic influence of vision on reaching activity by studying the interplay between visual feedback and arm movement-related signals during the execution of reaching movements performed towards targets located in different positions

(see activity in collaboration between UNIBO and UJI in Task 5.1 and the specific experimental tasks descrive in Section 2.2. The steering actions were undertaken to better characterize the dynamic properties of the medial parieto-occipital cortex as depicted in Task 5.1. In this line, the analysis of the interplay between reach-directions and retinotopic/spatial position of targets has been undertaken in collaboration between UJI and UNIBO (see synthesis in Task 5.2).
In addition, a new paradigm has been set and used to record neurons in the medial parieto-occipital cortex, so to highlight neural modulations reflecting the displacement of the spotlight of attention in absence of motor (oculomotor and arm) output (see activity in collaboration between UNIBO and WWU in Task 5.3).
These shaping of the activities conforms with a "tuning" of the experimental activity with the general objectives of the EYESHOTS Project in its executive phase.

4 Deliverables and milestones tables

Deliverables (excluding the periodic and final reports).

TABLE 1. DELIVERABLES

Del. no.	Deliverable name	WP no.	Lead beneficiary	Nature	Dissemination level	Delivery date from Annex I (proj month)	Delivered Yes/No	Actual / Forecast delivery date	Comments
D4.2a	Computational approach to the integration problem	WP4	UJI	R	PU	15	Yes	8-Jun-09 / 31-May-09	
D5.3a	Report on the respective influence of motor and visual parameters on fragment location obtained from saccade adaptation data on humans	WP5	WWU	R	PU(*)	15	Yes	5-Jun-09 / 31-May-09	
D1.2	Non-visual depth cues and their influence on perception	WP1	UG	R	PU	18	Yes	9-Sep-09 / 31-Aug-09	
D2.1	Convolutional network for vergence control	WP2	K.U.Leuven	R	PU	18	Yes	9-Sep-09 / 31-Aug-09	
D1.4	Bioinspired Stereovision Robot System	WP1	UG	P	PP	24	Yes	4-Mar-10 / 28-Feb-10	
D2.2a	Algorithm for 3D scene description through interactive visual stereopsis adaptation using a conventional binocular vision platform	WP2	K.U.Leuven	O	PU(*)	24	Yes	1-Mar-10 / 28-Feb-10	
D3.1b	Demonstration of object selective cells at intermediate complexity showing properties of disparity	WP3	WWU	R	PU	24	Yes	10-Mar-10 / 28-Feb-10	
D3.3a	Working memory model	WP3	WWU	O	PU(*)	24	Yes	1-Mar-10 / 28-Feb-10	
D5.1	Report on neural discharges in the medial parieto-occipital cortex	WP5	UNIBO	R	PU(*)	24	Yes	4-Mar-10 / 28-Feb-10	

(*) According to Annex I these deliverables will be made publicly available after the corresponding material will have been accepted for publication in journals/conf.proceedings

Milestones

TABLE 2. MILESTONES							
Milestone no.	**Milestone name**	**Work package no.**	**Lead beneficiary**	**Delivery date from Annex I**	**Achieved Yes/No**	**Actual / Forecast achievement date**	**Comments**
M5	Convolutional network algorithm for sparse near/far coding	WP2	K.U.Leuven	15	Yes	May 2009	The convolutional network has been demonstrated to the partners UG and WWU and at the general meeting in Milan, 13-14/1/2010. The software has been made available on the EYESHOTS repository for all partners to use. The milestone has been reached as planned.
M6	Monkey training for neural recording completed	WP5	UNIBO	15	Yes	May 2009	It consisted in performing the training of the monkey in order to be in the condition of starting the electrophysiological sessions. This milestone has been achieved, so the subsequent session of recording from neurons has been possible (see Deliverable 5.1) . Data from recorded neurons are the explicit verification that this Milestone was reached.

M7	Target location (saliency map) for the next eye movement in goal-directed search	WP3	WWU	22	Yes	December 2009	At milestone M7 the saliency map should indicate the location of a target object by an increased activity as compared to other values at locations of distractor objects. The milestone was reached as expected. The validity was demonstrated with different target objects. This model was superior to a model without goal directed, top- down signals.
M8.ante	Experimental data on fragment location in monkeys obtained	WP5	UNIBO	22	Yes	December 2009	It consisted in the collection of populations of neurons from the medial parieto-occipital cortex while the monkey performed various visuomotor, and oculomotor tasks. This milestone has been achieved and the electrophysiological data have been shared with the consortium, so the other partners had the possibility to verify it. A part of the data are now on the way to be published, see PRR II year.

M8	End of single cell recording session	WP5	UNIBO	24	Yes	February 2010	As an intermediate result, a preliminary analysis of neuronal data (on fix-in-depth and reach-in-depth tasks) is available to partners. Specific single cell recordings related to Task 5.3 and 5.4 (cooperation between UNIBO and WWU) are in progress
M9.ante	Experimental data in single actor setting obtained	WP5	WWU	18	Yes	August 2009	The milestone presented an intermediate result. and was reached as planned. Preliminary data on the allocation of attention and behaviour prediction in humans and monkeys was made available to EYESHOTS partners. The hypothesis that gaze tracking can be used to predict allocation of attention and behaviour was confirmed by the experiment.

9 References

Adelson, E. H., and Bergen, J. R., The Plenoptic Function and the Elements of Early Vision. In: *Computational Models of Visual Processing.* In: M. Landy and J. A. Movshon, Cambridge, MA: MIT Press, 1991.

Antonelli M., Chinellato E., del Pobil, A.P. Visuomotor spatial awareness through concurrent reach/gaze actions. *CogSys 2010*, Zurich, Jaunuary 27-28, 2010.

Anzai, A., Ohzawa, I. and Freeman, R. D. Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *J Neurophysiol*, **82**(2):874-890, 1999.

Anzai, A., Ohzawa, I. and Freeman, R. D. Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proc Natl Acad Sci U S A*, **94**(10):5438-5443, 1997.

Awater, H., Burr, D., Lappe, M., Morrone, M. C. and Goldberg, M. E. The effect of saccadic adaptation on the localization of visual targets. *J. Neurophysiol*., **93**:3605-3614, 2005.

Bahcall, D. O. and Kowler, E. Illusory shifts in visual direction accompany adaptation of saccadic eye movements. *Nature*. **400:**864-66, 1999.

Barlow, H. B., Blakemore, C. and Pettigrew, J. D. The neural mechanism of binocular depth discrimination. *J Physiol*, **193**(2):327-342, 1967.

Berton, F., Sandini, G. and Metta G., Anthropomorphic visual sensors. *Encyclopedia of Sensors*, American Scientific Publishers, **10**:1-16, 2006.

Bichot, N.P., Rossi, A.F. and Desimone, R. Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, **308**:529-534, 2005.

Bisley, J.W. and Goldberg, M.E. Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, **299**:81-86, 2003.

Blakemore, C., Fiorentini, A., and Maffei, L. A second neural mechanism of binocular depth discrimination. *J Physiol*, **226**(3):725-749, 1972.

Blohm, G., Khan, A.Z., Ren, L., Schreiber, K.M., and Crawford, J.D. Depth estimation from retinal disparity requires eye and head orientation signals. *JOV*, **8**(16):3, 1-23, 2008.

Bolduc, M. and Levine, M. D. A Review of Biologically Motivated Space-Variant Data Reduction Models for Robotic Vision. *Computer Vision and Image Understanding*, **69**(2):170-184, 1998.

Bosco, A., Breveglieri, R., Chinellato, E., Galletti, C. and Fattori, P. Influence of visual feedback on reaching activity in parietal area V6A. *Program No. 355.20/Z1. 2009 Neuroscience Meeting Planner. Chicago, IL: Society for Neuroscience,* 2009. Online.

Bosco, A., Breveglieri, R., Chinellato, E., Galletti, C. and Fattori, P. Complex modulation of reaching activity by visual feedback in parietal area V6A. In preparation.

Brain, W.R. Visual disorientation with special reference to lesions of the right cerebral hemisphere. *Brain*, **64**:244-272, 1941.

Breazeal, C., Kidd, C., Thomaz, A.L., Hoffman, G., and Berlin, M. Effects of Nonverbal Communication on Efficiency and Robustness in Human-Robot Teamwork. In Proceedings of the *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005.

Breveglieri, R., Bosco, A., Canessa, A., Fattori, P. and Sabatini, S. P. Evidence for Peak-shaped Gaze Fields in Area V6A: Implications for Sensorimotor Transformations in Reaching Tasks. In*: Bioinspired Applications in Artificial and Natural Computation*, **2**: pages 324-333, Springer-Verlag, Berlin-Heidelberg, 2009a.

Breveglieri, R., Fattori, P., Bosco, A., Lappe, M. and Galletti, C. Covert spatial displacements of the spotlight of attention modulates neuronal activity in area V6A of the medial parieto-occipital cortex of the monkey. *Program No. 758.16/AA15. 2009 Neuroscience Meeting Planner. Chicago, IL: Society for Neuroscience, 2009. Online*. 2009b.

Bridge, H. and Cumming, B.G. Responses of macaque v1 neurons to binocular orientation differences. *J Neurosci.* **21**(18):7293-7302, 2001.

Canessa A., Sabatini S.P., and Solari F. Visuo-motor constraints in binocular eye coordination: optimization theories revisited. *JOV*, submitted, 2010.

Capurro, C., Panerai, F., and Sandini, G. Dynamic vergence using log-polar images. *Int. J. Comp. Vis*., **24**(1):79-94, 1997.

Castiello, U. Grasping a fruit: selection for action. *J. Exp. Psychol. Hum. Percept. Perform*, **22**:582-603 , 1996.

Cavanaugh, J. and Wurtz, R.H. Subcortical modulation of attention counters change blindness. *J. Neurosci*., **24**:11236-11243., 2004.

Chang, J.Y., Chen, L., Luo, F., Shi, L.H. and Woodward, D.J. Neuronal responses in the frontal cortico-basal ganglia system during delayed matching-to-sample task: ensemble recording in freely moving rats. *Exp Brain Res.*, **142**:67-80, 2002.

Chelazzi, L, Miller, EK, Duncan, J., Desimone, R. A neural basis for visual search in inferior temporal cortex. *Nature*, **363**: 345-347, 1993.

Chessa, M., Sabatini, S.P., and Solari, F. A fast joint bioinspired algorithm for optic flow and two-dimensional disparity estimation. In ICVS, pages 184-193, 2009a.

Chessa, M., Sabatini, S. and Solari, F., A Virtual Reality Simulator for Active Stereo Vision Systems, in: VISSAPP, **2**, pages 444-449, 2009b.

Chessa M., Canessa A., Gibaldi A., Solari F., and Sabatini S.P. Embedding fixation constraints into binocular energy-based models of depth perception. *ICCNS*, Boston, MA, USA, May 27-30, 2009c.

Chinellato, E. and del Pobil, A.P. Neural Coding in the Dorsal Visual Stream. From Animals to Animats, In: *International Conference on the Simulation of Adaptive Behavior*, pages 230-239, 2008.

Chinellato, E., Grzyb, B. J., Marzocchi, N., Bosco, A., Fattori, P. and del Pobil, A. P. Eye-Hand Coordination for Reaching in Dorsal Stream Area V6A: Computational Lessons., In*: Bioinspired Applications in Artificial and Natural Computation*, **2**:304-313. Springer-Verlag, Berlin-Heidelberg, 2009a.

Chinellato, E., Grzyb1, B. J., Fattori, P. and del Pobil, A. P. Toward an Integrated Visuomotor Representation of the Peripersonal Space., In*: Bioinspired Applications in Artificial and Natural Computation*, **2**:304-313. Springer-Verlag, Berlin-Heidelberg, 2009b.

Chinellato E., Antonelli, M., del Pobil, A. Visuomotor spatial awareness through concurrent reach/gaze actions. Accepted in: *Transactions on Autonomous Mental Development*, 2010.

Chumerin, N., Gibaldi, A., Sabatini, S.P., and Van Hulle, M.M. Convolutional Network for Vergence Control. *ISABEL 2009,* Bratislava, Slovak Republic, November 24-27, pages 1-6, 2009.

Chumerin, N., Gibaldi, A., Sabatini, S.P., and Van Hulle, M.M. Learning eye vergence control from a distributed disparity representation. *International Journal of Neural Systems*, accepted with revision, 2010.

Coello, Y. Spatial context and visual perception for action. *Psicologica*, **26:**39-59, 2005.

Coello, Y., Bartolo, A., Amiri, B., Devanne, H., Houdayer, E. and Derambure, P. Perceiving what is reachable depends on motor representations: evidence from a transcranial magnetic stimulation study. *PLoSOne*, **3**(8):e2862, 2008.

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

Collins, T., Rolfs, M., Deubel, H., and Cavanagh, P. Post-saccadic location judgments reveal remapping of saccade targets to non-foveal lo-cations. *JOV*, **9**(5):1-9, 2009.

Dang, T., Hoffmann, C. and Stiller, C. Continuous Stereo Self-Calibration by Camera Parameter Tracking. IEEE Transactions on Image Processing, **18**(7):1536-1550, 2009.

De Valois R.L., Albrecht D.G., and Thorell L.G. Spatial frequency selectivity of cells in macaque visual cortex. *Vis. Res*, **22**(5):545-559, 1982.

Deubel, H., Schneider, W.X. Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vis. Res*, **36**:1827-1837, 1996.

Deubel, H., Schneider, W.X., Paproppa, I. Selective dorsal and ventral processing: Evidence for a common attentional mechanism in reaching and perception. *Vis. Cogn*., **5**:81-107, 1998.

Fattori P and Galletti C. Neuronal processes of action in the superior parietal lobule of primate brain. *Frontiers in Neuroscience. Conference Abstract: 3rd Mediterranean Conference of Neuroscience,* doi: 10.3389/conf.neuro.01.2009.15.025, 2009.

Fattori, P., Bosco, A., Breveglieri, R., Marzocchi, N., and Galletti, C. Visual and somatosensory guidance of reaching movements in the medial parieto-occipital cortex of the macaque. *JOV*, **9**(8):697a, 2009.

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

Fattori, P., Kutz, D.F., Breveglieri, R., Marzocchi, N. and Galletti, C. Spatial tuning of reaching activity in the medial parieto-occipital cortex (area V6A) of macaque monkey. *Eur J Neurosci*, **22**:956-972, 2005.

Ferster, D. A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. *J. Physiol*, **311**:623-655, 1981.

Fischl, B., Cohen, M., E. Schwartz, The local structure of space-variant images, Neural Networks, **10**(5):815-831, 1997.

Fitzpatrick, P., Metta, G., Natale, L., Rao, S., and Sandini, G. Learning about objects through action - initial steps towards artificial cognition. In: *IEEE Intl. Conf. on Robotics and Automation (ICRA),* Taipei, Taiwan, May 12 - 17 2003.

Florack, L. Modeling Foveal Vision, in: *Scale Space and Variational Methods in Computer Vision*, pages 919-928, 2007.

Franz, A. and Triesch, J. Emergence of disparity tuning during the development of vergence eye movement, In: *Proc. Int. Conf. on Development and Learning*, pages 31-36. Piscataway, NJ:IEEE, 2007.

Fuke, S., Ogino, M., and Asada, M. Acquisition of the Head-Centered Peri-Personal Spatial Representation Found in VIP Neuron. *IEEE Transactions on Autonomous Mental Development*, **1**:131-140, 2009.

Galletti, C., Battaglini, P.P. and Fattori, P. Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. *Eur J Neurosci*.,**7**:2486-2501, 1995.

Galletti, C., Breveglieri, R., Lappe, M., Bosco, A., Ciavarro, M. and Fattori, P. Covert shift of attention modulates the ongoing neural activity in a reaching area of the macaque dorsomedial visual stream Submitted, 2010.

Galletti, C., Fattori, P., Battaglini, P.P., Shipp, S. and Zeki, S. Functional demarcation of a border between areas V6 and V6A in the superior parietal gyrus of the macaque monkey. *Eur J Neurosc.*, **8**:30-52, 1996.

Galletti, C., Fattori, P., Kutz, D.F. and Battaglini, P.P. Arm movement-related neurons in the visual area V6A of the macaque superior parietal lobule. *Eur J Neurosci*., **9**:410-413, 1997.

Galletti, C., Fattori, P., Kutz, D.F. and Gamberini, M. Brain location and visual topography of cortical area V6A in the macaque monkey. *Eur J Neurosc.*, **11**:575-582, 1999.

Galletti, C., Kutz, D.F., Gamberini, M., Breveglieri, R. and Fattori, P. Role of the medial parieto-occipital cortex in the control of reaching and grasping movements. *Exp Brain Res.*, **153**:158-170, 2003.

Gamberini, M., Passarelli, L., Fattori, P., Zucchelli, M., Bakola, S., Luppino, G. and Galletti, C. Cortical connections of the visuomotor parietooccipital area V6Ad of the macaque monkey. *J Comp Neurol*, **513**:622-642, 2009.

Genovesio, A. and Ferraina, S. Integration of retinal disparity and fixation-distance related signals toward an egocentric coding of distance in the posterior parietal cortex of primates. *J Neurophysiol*., **91**:2670-2684. Epub 2004 Feb 2611, 2004.

Gibaldi A., Chessa M., Canessa A., Sabatini S.P., and Solari F. A cortical model for binocular vergence control without explicit calculation of disparity. *Neurocomp.,* **73**:1065-1073, 2010a.

Gibaldi A., Canessa A., Chessa M., Sabatini S.P., and Solari F. Read-out rules for short-latency disparity-vergence responses from populations of binocular energy units: the effect of vertical disparities. Submitted to *ECVP'10*, Lausanne, 22-26 August, 2010b.

Gnadt. J.W. and Mays, L.E. Neurons in monkey parietal area LIP are tuned for eye-movement parameters in three-dimensional space. *J Neurophysiol.,* **73**:280, 1995.

Greenwald, H. S., and Knill, D. C. Cue integration outside central fixation: A study of grasping in depth. *JOV*, **9**(2):11, 1-16, 2009.

Hamker, F.H. The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral Cortex.*, **15**:431-447, 2005.

Hamker, F. H., Zirnsak, M. V4 receptive field dynamics as predicted by a systems-level model of visual attention using feedback from the frontal eye field. *Neural Networks*, **19**:1371-1382, 2006.

Hamker, F. H. Modeling feature-based attention as an active top-down inference process. *BioSystems*, **86**:91-99, 2006.

Hamker, F. H. The mechanisms of feature inheritance as predicted by a systems-level model of visual attention and decision making. *Advances in Cognitive Psychology*, **3**:111-123, 2007.

Hamker, F.H., Zirnsak, M., Calow, D., Lappe, M. The peri-saccadic perception of objects and space. *PLOS Computational Biology*, **4(2)**:e31, 2008.

Hansard, M. and Horaud, R. Cyclopean geometry of binocular vision. *JOSA A,* **25**:9, 2357, 2008.

Hansard, M. and Horaud, R. Cyclorotation Models for Eyes and Cameras. *IEEE Trans. on System, Man, and Cybernetics--Part B: Cybernetics*, **40**(1):151-161, 2010.

Hayhoe, M. and Ballard, D. Eye movements in natural behavior. *Trends in Cognitive Science,* **9**(4):188-94, 2005.

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

Henderson, J. M. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences,* **7:**498-504, 2003.

Henderson, J. M., and Hollingworth, A. Eye movements and visual memory: detecting changes to saccade targets in scenes. *Perception & Psychophysics*, **65:**58-71, 2003.

Hoffman, J.E. and Subramaniam, B. The role of visual attention in saccadic eye movements. *Percept Psychophys*, **57**:787-795, 1995.

Holmes, G. Disturbance of visual space perception. *Br Med J,* **2**:230-233, 1919.

Holmes, G., Horrax, G. Disturbances of spatial orientation and visual attention with loss of steroscopic vision. *Archives of Neurology and Psychiatry*, **1**:385-407, 1919.

Horng, J, Semmlow, J., Hung, G.K., and Ciuffreda, K. Initial Component in Disparity Vergence: A Model-Based Study. *IEEE Trans. Biomed. Eng.*, **45**(2): 249-257, 1998.

Howard, I. P., and Rogers, B. J. *Seeing in depth: Volume 2, Depth perception*. Ontario, Canada: I Porteous Publishing, 2002.

Hung, G.K., Semmlow, J.L., and Ciuffreda, K.J. A dual-mode dynamic model of the vergence eye movement system. *IEEE Trans. Biomed. Eng.*, **36**(11):1021-1028, 1986.

Hutchinson, S., Hager, G.D., and Corke, P.I. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, **12**(5):651-670, 1996.

Ignashchenkova, A., Dicke, P.W., Haarmeier, T., Thier, P. Neuron-specific contribution of the superior colliculus to overt and covert shifts of attention. *Nat Neurosci*., **7**:56-64, 2004.

Jamone, L. , Metta, G., Nori, F., and Sandini, G. James, a humanoid robot acting over an unstructured world. In*: Proc. of the Humanoids 2006 conference*, December 4–6th, 2006, Genoa, Italy, 2006.

Jampel, R.S. The function of the extraocular muscles, the theory of the coplanarity of the fixation planes*. J. Neurolog. Sci.,* **280:**1-9, 2008.

Joel, D., Niv, Y., and Ruppin, E. Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw*., **15**(4-6):535-47, 2002.

Jonides, J., Irwin, D.E. and Yantis, S. Integrating visual information from successive fixations. *Science*, **215**:192-4, 1982.

Jonikaitis, D., Deubel, H. Split attention during simultaneous eye and hand movements. *Perception ECVP Abstract Supplement*, **38**:158, 2009.

Joshua, D. E. and Bishop, P. O. Binocular single vision and depth discrimination: receptive field disparities for central and peripheral vision and binocular interaction on peripheral single units in cat striate cortex. *Exp Brain Res*., **10**(4):389-416, 1970.

Jurie, F. A new log-polar mapping for space variant imaging. - Application to face detection and tracking, *Pattern Recognition*, **32**: 865-875, 1999.

Kowler, E., Anderson, E., Dosher, B., Blaser, E. The role of attention in the programming of saccades. *Vis. Res.*, **35**:1897, 1995.

Krishnan, V.V., and Stark, L. A heuristic model for the human vergence eye movement system, *IEEE Trans Biomed Eng*, **24**:44-9, 1977.

Kumar, S., Behera, L., and McGinnity, T. M. Kinematic control of a redundant manipulator using an inverse-forward adaptive scheme with a KSOM based hint generator. *Robot. Auton. Syst.,* **58(**5):622-633, 2010.

Kutz, D.F., Fattori, P., Gamberini, M., Breveglieri, R. and Galletti, C. Early- and late-responding cells to saccadic eye movements in the cortical area V6A of macaque monkey. *Exp Brain Res*., **149**:83-95, 2003.

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

LeVay, S. and Voigt, T. Ocular dominance and disparity coding in cat visual cortex. *Vis Neurosci*., **1**(4):395-414, 1988.

Manzotti, R., Gasteratos, A., Metta, G. and Sandini, G. Disparity Estimation on Log-Polar Images and Vergence Control, *Computer Vision and Image Understanding*, **83** (2), 97-117, 2001.

Martinez-Trujillo, J.C. and Treue, S. Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol*., **14**: 744-751, 2004.

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

McLaughlin, S.C. Parametric adjustment in saccadic eye movements. *Percept Psychophys*, 2:359362, 1967.

Miller, E., Cohen, J. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*., **24**:167-202, 2001.

Mok, D., Ro, A., Cadera, W., Crawford, J.D., and Vilis., T. Rotation of listing's plane during vergence. *Vision Res.*, **32**:2055–2064, 1992.

Monaco, J.P., Bovik, A.C., Cormack, L.K. Active, foveated, uncalibrated stereovision. *Int. J. of Computer Vision,* **85**(2):192-207, 2009.

Moore, T., Armstrong, K.M. Selective gating of visual signals by microstimulation of frontal cortex. *Nature***, 421**:370-373, 2003.

Munuera, J., Morel, P., Duhamel, J., and Deneve, S. Optimal sensori-motor control in eye movement sequences. *J. Neurosci*., **29**(10):3026-3035, 2009.

Mutlu, B.**,** Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. Footing in Human-Robot Conversations: How Robots Might Shape Participant Roles Using Gaze Cues. In: *Proceedings of the 4th ACM/IEEE Conference on Human-Robot Interaction*, 2009b.

Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., and Hagita, N. Nonverbal Leakage in Robots: Communication of Intentions through Seemingly Unintentional Behavior. In: *Proceedings of the 4th International Conference on Human-Robot Interaction* , 2009a.

Nakamura, K., Chung, H.H., Graziano, M.S.A. and Gross, C.G. Dynamic representation of eye position in the parieto-occipital sulcus. *J Neurophysio*l., **81**:2374-2385, 1999.

Natale, L., Nori, F., and Metta, G. Learning precise 3d reaching in a humanoid robot. In: *Proc. 6th IEEE Int. Conf. Develop. Learn.*, 2007.

Naya, Y., Yoshida, M., Takeda, M., Fujimichi, R. and Miyashita, Y. Delay-period activities in two subdivisions of monkey inferotemporal cortex during pair association memory task. *Eur J Neurosci*., **18**:2915-2918, 2003.

Nelson, J.I., Kato, H., Bishop, P.O. Discrimination of orientation and position disparities by binocularly activated neurons in cat striate cortex. *J Neurophysiol*., **40**(2):260-283, 1977.

Nikara, T., Bishop, P. O. and Pettigrew, J. D. Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp Brain Res*., **6**(4):353-372, 1968.

O'Regan, J.K. and Levy-Schoen, A. Integrating visual information from successive fixations: does trans-saccadic fusion exist? *Vis. Res.,* **23**:765-8, 1983.

O'Regan, J. K., and Noe, A. A sensorimotor account of vision and visual consciousness. *Behavioral Brain Science*, 24, 939-1031, 2001.

O'Reilly, R. C. and Frank, M. J. Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput*, **18**(2):283-328, 2006.

Panouilleres, M., Weiss, T., Urquizar, C., Salemme, R., Munoz, D. P., and Pelisson, D. Behavioral evidence of separate adaptation mechanisms controlling saccade amplitude lengthening and shortening. *J. Neurophysiol.*, **101**(3):1550-1559, 2009.

Patel S.S., Ogmen H., and Jiang B.C.. Neural network model of short-term horizontal disparity vergence dynamics. *Vis. Res*., **37**(10):1383-1399, 1996.

Pauwels, K. and Van Hulle, M.M. Realtime phase-based optical flow on the gpu. In *CVPR Workshop on Computer Vision on the GPU*, 2008

Perenin, M.T. and Vighetto, A. Optic ataxia: a specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain*, **111**:643-674, 1988.

Pobuda, M. and Erkelens, C.J. The relationship between absolute disparity and ocular vergence. *Biolog. Cybern.*, **68**(3):221-228, 1993.

Poggio, G. E. Mechanisms of stereopsis in monkey visual cortex. *Cereb Cortex*., **5**(3):193-204, 1995.

Poggio, G. F. and Fischer, B. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J Neurophysiol*. **40**(6):1392-1405, 1977.

Poggio, G. F. and Talbot, W. H. Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J Physiol*., **315**:469-492, 1981.

Poggio, G. F., Gonzalez, F. and Krause, F. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J Neurosci*., **8**(12):4531-4550, 1988.

Pouget, A., and Sejnowski, T.J. Spatial Transformations in the Parietal Cortex Using Basis Functions. *J. Cog. Neurosci*. **9**(2):222-237, 1997.

Prince, S.J D, Cumming, B. G. and Parker, A. J. Range and mechanism of encoding of horizontal disparity in macaque V1. *J Neurophysiol*., **87**(1):209-221, 2002.

Rambold, H.A. and Miles, F.A. Human vergence eye movements to oblique disparity stimuli: evidence for an anisotropy favoring horizontal disparities. *Vis. Res*., **48**(19):2006-2019, 2008.

Ranganath, C. Working memory for visual objects: complementary roles of inferior temporal, medial temporal, and prefrontal cortex. *J. Neurosci*., **139**:277-289, 2006.

Rashbass, C. and Westheimer, G. Disjunctive Eye Movements. *J. Phisyol*., **159**:339-360, 1961.

Rasolzadeh, B., Bjorkman, M., Huebner, K., and Kragic D. An active vision system for detecting, fixating and manipulating objects in real world. *IJRR*, **27**(2-3): 133-154, 2010.

Read, J.C.A. and Cumming, B.G. Understanding the cortical specialization for horizontal disparity. *Neural Computation*, **16**(10):1983-2020, 2004.

Read, J.C.A., and Cumming, B.G. Does depth perception require vertical-disparity detectors? *JOV*, **6**(12):1, 1323-1355, 2006.

Read, J.C.A., Phillipson, G. P., and Glennerster, A. Latitude and longitude vertical disparities. *JOV*, **9**(13):11, 1-37, 2009.

Reynolds, J.H,, Chelazzi, L. Attentional modulation of visual processing. *Annu. Rev. Neurosci.,* **27**:611-647, 2004.

Rizzolatti G., Ferrari P.F., Rozzi S., and Fogassi L. The inferior parietal lobule: where action becomes perception. *Novartis Found. Symp.,* **270:** 129-40, 2006.

Rougeaux, S., and Kuniyoshi, Y. Robust Tracking by a Humanoid Vision System. In: *IARP First International*

Workshop on Humanoid and Human Friendly Robotics. Tsukuba Science City, Japan, 1998.

Rousselet, G.A., Thorpe, S.J. and Fabre-Thorpe, M. How parallel is visual processing in the ventral pathway? *Tr. Cog. Sci.,* **8**:363-70, 2004.

Sabatini S.P., Gastaldi G., Solari F., Pauwels K., Van Hulle M.M., Diaz J., Ros E., Pugeault N., and Kruger N.. A compact harmonic code for early vision based on anisotropic frequency channels. Computer Vision and Image Understanding, in press

Sakata, H., Shibutani, H., Kawano, K. Spatial properties of visual fixation neurons in posterior parietal association cortex of the monkey. *J. Neurophysiol.,* **43**:1654-1672, 1980.

Salinas, E. and Their, P. Gain modulation: a major computational principle of the central nervous system. *Neuron*. **27**:15-21, 2000.

Santini, F., and Rucci, M. Active estimation of distance in a robotic system that replicates human eye movement. *Robotics and Autonomous Systems*. **55**(2):107-121, 2007.

Saxena, A., Driemeyer, J., Ng, A. Y. Robotic Grasping of Novel Objects using Vision. *IJRR*, **27**(2):157, 2008.

Schindler, K., Geometry and construction of straight lines in log-polar images, *Computer Vision and Image Understanding*, **103**(3): 196-207, 2006.

Schor, C. M., Maxwell, J. S., McCandless, J., & Graf, E. . Adaptive control of vergence in humans. *Annals of the New York Academy of Sciences,* **956**:297-305, 2002.

Schor, C.M.. The relationship between fusional vergence eye movements and fixation disparity. *Vis. Res.,* **19**(12):1359-1367, 1979.

Schreiber, K., Crawford, J.D., Fetter, M. and Tweed, D. The motor side of depth vision. *Nature*, **410**:819-822, 2001.

Schreiber, K.M., Tweed, D.B. and Schor, M.C. The extended horopter: quantifying retinal correspondence across changes of 3D eye position. *JOV*, **6**:64-74, 2006.

Schreiber, K. M., Hillis, J. M., Filippini, H. R., Schor, C. M., & Banks, M. S. (2008). The surface of the empirical horopter. *Journal of Vision*, **8**(3):7, 1-20

Schwartz, E.L.. Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biolo. Cyber.,* **25**(4):181-194, 1977.

Serrano-Pedraza, I. and Read, J.C.A. Stereo vision requires an explicit encoding of vertical disparity. *JOV,* **9**(4):1-13, 2009.

Serrano-Pedraza, I., Phillipson, G. P., and Read, J. C. A. A specialization for vertical disparity discontinuities. *Journal of Vision*, **10**(3):2, 1-25, 2010.

Serre, T., et al. Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.*, **29**:411-426, 2007.

Sheliga, B. M., & Miles, F. A. Perception can influence the vergence responses associated with open-loop gaze shifts in 3D. *Journal of Vision*, **3**(11):654-676, 2003.

Solari F., Chessa M., and Sabatini S.P. Design strategies for direct multiscale and multi-orientation visual processing in the log-polar domain. *Pattern Recognition Letters*, submitted.

Staudte, M., and Crocker, M.W. The utility of gaze in spoken human-robot interaction. In: *Proceedings of "Metrics for Human-Robot Interaction",* Workshop at ACM/IEEE HRI, 2008.

Theimer, W.M., and Mallot, H.A. Phase-based vergence control and depth reconstruction using active vision. *CVGIP*, Image understanding, **60**(3):343-358, 1994.

Thompson, K.G., Biscoe, K.L., Sato, T.R. Neuronal basis of covert spatial attention in the frontal eye field. *J. Neurosci.* **25**:9479-9487, 2005.

Traver, V., and Pla, F. Log-polar mapping template design: From task-level requirements to geometry parameters, *Image Vision and Computing*, **26**(10):1354-1370, 2008.

Traver, V., Bernardino, A. A review of log-polar imaging for visual perception in robotics. *Robotics and Autonomous Systems*, **58**(4): 378-398, 2010.

Treue, S., Martinez-Trujillo, J.C. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature,* **399**:575-579, 1999.

Tweed, D., and Vilis, T. Geometric relations of eye position and velocity vectors during saccades. *Vis. Res.*, **30**(1):111–127, 1990.

Tweed, D. Visual-motor optimization in binocular control. *Vis. Res.,* **37:**1939-1951, 1997.

Varela, F. J., Thompson, E. T., and Rosch, E. The embodied mind. *MIT Press.* 1992.

Vitay, J. and Hamker, F. H. Sustained activities and retrieval in a computational model of the perirhinal cortex. *J. Cogn. Neurosci.*, **20**(11):1993-2005, 2008.

Vitay, J., Hamker, F. H. A computational model of the influence of basal ganglia on memory retrieval in rewarded visual memory tasks. *Frontiers of Comp. Neurosci.*, under revision.

Vitay, J., Fix, J., Beuth, F., Schroll, and H., Hamker, F.H. Biological Models of Reinforcement Learning. *Künstliche Intelligenz*, **3**:12-18, 2009.

von der Heydt, R., Adorjani, C., Hänny and P., Baumgartner, G. Disparity sensitivity and receptive field incongruity of units in the cat striate cortex. *Exp. Brain Res.,* **31**(4):523-545, 1978.

Wan, D., and Zhou, J. Self-calibration of spherical rectification for a PTZ-stereo system. *Image and Vision Computing*, **28**:367-375, 2010.

Wang, Y., and Shi, B.E. Autonomous development of vergence control driven by disparity energy neuron populations. *Neural Comp,* **22**:1-22, 2009.

Watanabe, Y., Funahashi, S. Neuronal activity throughout the primate mediodor- sal nucleus of the thalamus during oculomotor delayed-responses. I. Cue-, delay-, and response-period activity. *J. Neurophysiol.*, **92**:1738-1755, 2004.

Wei, K. and Kording, K. Relevance of error: What drives motor adaptation? *J. Neurophysiol.*, **101**(2):655-664, 2009.

Westheimer, G., and Mitchell, A.M. Eye movement responses to convergence stimuli. *Archives of Ophthalmology*, **55**(6):848, 1956.

Wiltschut, J., and Hamker, F. Efficient coding correlates with spatial frequency tuning in a model of v1 receptive field organization. *Vis. Neurosci*., **26**:21-34, 2009.

Wörgötter F., Porr B. Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms. *Neural Comp.*, **17**:245-319, 2005.

Yang, D.S., FitzGibbon , E.J. and Miles, F.A. Short-latency disparity-vergence eye movements in humans: sensitivity to simulated orthogonal tropias. *Vision Research* **43**:431-443, 2003.

Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., and Miyamoto T. The effects of responsive eye movement and blinking behavior in a communication robot. In: *Proceedings of IEEE/RSJ International Coference on Intelligent Robots and Systems*, pages 4564-4569, 2006.

Zhang, P.Y., Lu, T.S., and Song, L.B. RBF networks-based inverse kinematics of 6R manipulator. In: *International Journal of Advanced Manufacturing Technology*, **26**(1-2): 144-147, 2005.

Zimmermann, E. and Lappe, M. Mislocalization of flashed and stationary visual stimuli after adaptation of reactive and scanning saccades. *J. Neurosci.*, **29**(35):11055-11064, 2009.

Zimmermann, E., and Lappe, M. Motor signals in visual localization. *JOV*, in press.

Zirnsak , M., Lappe, M., Hamker, FH. The spatial distribution of receptive field changes in a model of perisaccadic perception: Predictive remapping and shifts towards the saccade target. *Vis. Res.*, in press.

Zirnsak, M., Hamker, F.H. Attention alters feature space in motion perception. *J. Neurosci.*, in press.