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 01/03/2010 to 28/02/2011

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

prof. Silvio P. Sabatini

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	**3**
1. Publishable summary	**4**
1.1. Project's goal	4
1.2 Specific objectives	5
1.3 Expected final results	5
1.4 Work performed and main results achieved in the reporting period	5
2. Project objectives for the period	**8**
2.1 Overview	8
2.2 Follow-up of previous review	11
3. Work progress and achievements during the period	**15**
3.1 Progress overview and contribution to the research field	15
3.2 Workpackage progress	33
WP1 – Eye movements for exploration of the 3D space	39
WP2 – Active stereopsis	48
WP3 – Selecting and binding visual fragments	59
WP4 – Sensorimotor integration	65
WP5 – Human behavior and neural correlates of multisensory 3D representation	73
4. Deliverables and milestone tables	**90**
5. Project management	**94**
5.1 Management activities	94
5.2 Dissemination and use of the knowledge	97
6. Explanation of the use of the resources	**102**
6.1 Justification of major costs and resources	102
6.2 Budgeted versus actual costs	107
6.3 Planned versus actual effort	110
7. Financial statements – Form C and summary financial report	**111**
8. Certificates on the financial statements	**122**
9. References	**123**

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 2011

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 has run for a total of 36 months from the 1st of March 2008 to the 28th of February 2011. 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).

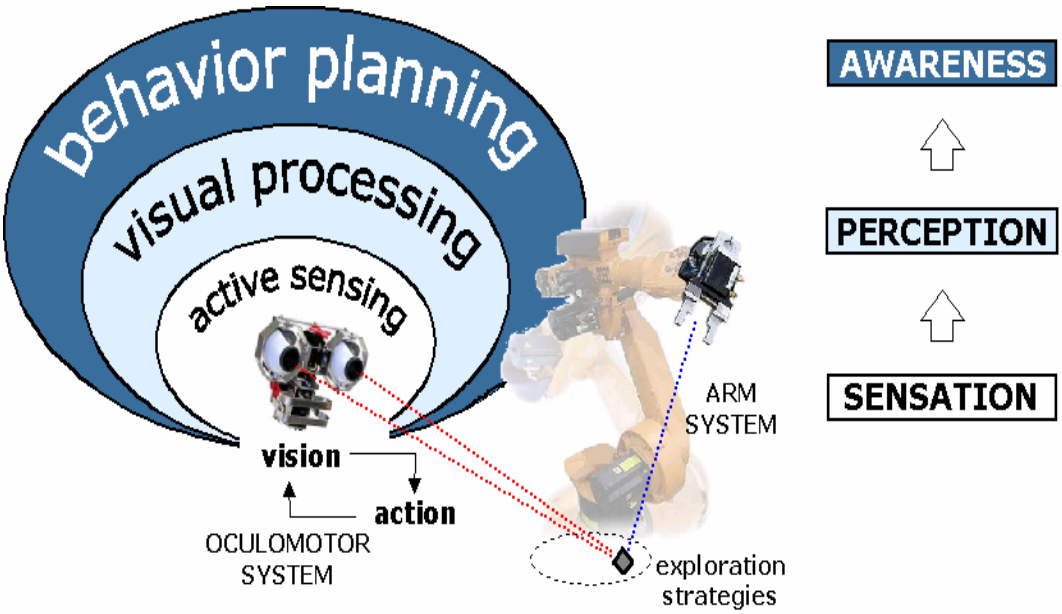


Figure 1: *The EYESHOTS perspective for 3D space perception.*

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 (01/03/10 – 28/02/11)

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** (February 2010) 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) 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.

In the third period, the work focused on the integration of the different modules on the robot platforms, in order to validate the approach in real-world conditions. Specific achievements are:

- The development of working modules for stereopsis and oculomotor control
  - *Distributed disparity energy model* [UG]
  - *Binocular object detection (and FEF saliency map)* [WWU]
  - *Vergence control modules* [K.U.Leuven, UG]
- The validation of interactive stereopsis behaviour in real-world situations on the i-Cub platform
  - *Dual-mode vergence control* [UG]
  - *Concurrent disparity and gaze estimation* [K.U.Leuven]
- The realization of the non-conventional tendon-driven mechatronic binocular system [UG]
- The integration of perceptual/visuomotor strategies on the eye/head-arm UJI robot platform.

A specific experimental framework has been devised to capture the *bilateral* interaction between motor and perceptual processes. In particular, we validated the benefit of coordinated execution/planning of binocular saccades and reaching actions (see Fig. 2) when a system is engaged in gazing and/or reaching foveated/not-foveated targets that are selected among multiple alternatives in the robot's peripersonal space.

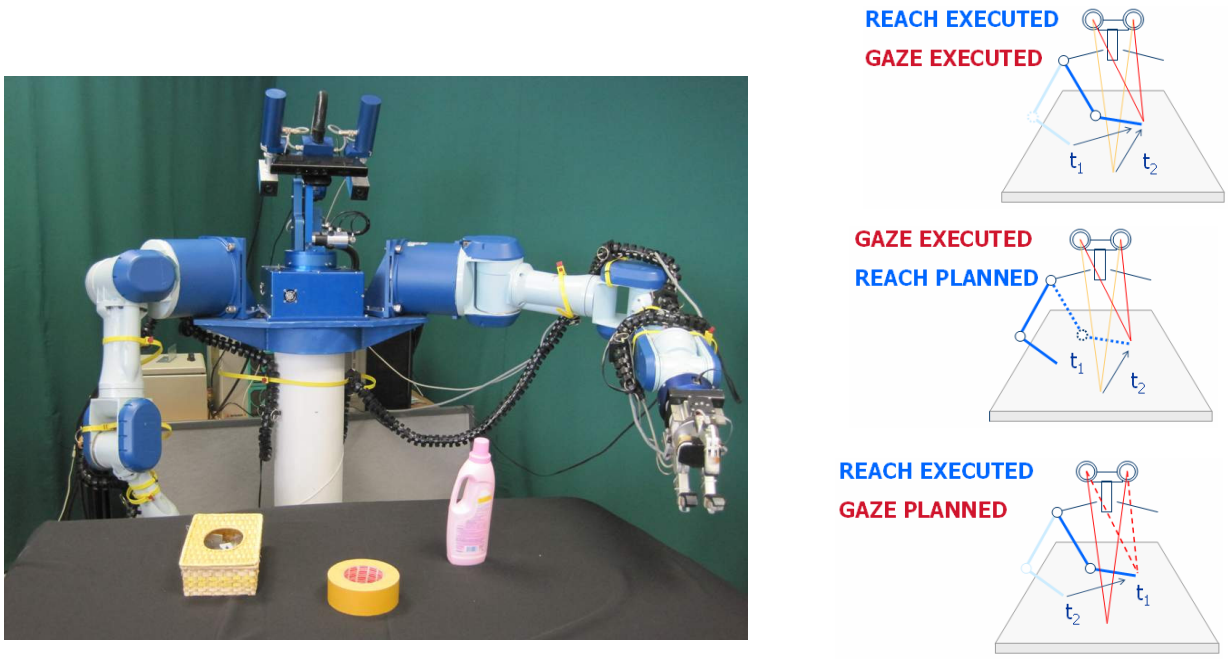


Figure 2: ***(Left)*** *The multi-object experimental setup used with the UJI humanoid robot* Tombatossals. ***(Right)*** *Concurrent gaze and reaching actions used to test the bilateral interaction between binocular oculomotor coordination and reaching tasks: saccades and reaching actions alternate sequentially both in the execution or planning phases. The robot employs the egocentric representation of peripersonal space it has gained, to interact with surrounding objects, recognize them and perform custom visuomotor and arm-motor actions, such as: foveate on the hand; reach the gazing point; show memory; foveate on a given object (either inside or outside the field of view); reach a given object (either foveated or not); execute a sequence of saccades by employing either covert or overt attention..*

Together with these integration activities, the ongoing development of models, as well the analysis of the data collected in neurophysiological and psychophysical experiments yielded significant results on the neuronal mechanisms used to link different fragments by the use of visual, attentional, oculomotor, and arm-movement related cues. 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 an integrated representation of visual, arm and eye sensorimotor information on which to base the 3D location of potential visual targets with respect to the body. The advantages from such representation, based on a Radial Basis Function framework, have been analyzed in theory, and have allowed us to implement advanced behaviors on the UJI humanoid robot setup.

Concerning the motor influence of visual perception, the analysis of the experimental data collected in the previous periods on saccadic adaptation by WWU is now completed. The analysis has yielded remarkable evidences of the oculomotor components of visual target localization, which have been included in the final model. Saccadic adaptation experiments were performed in simulation on the UJI model and on the real robot, providing interesting practical and theoretical insights.

The final status of the project is depicted in the diagram of Figure 3 evidencing the different project's components and their integration. The large greyed box represents the *EYESHOTS' Agent* engaged in the perception action cycle: The information flows circularly from the environment to sensory structures, to motor structures, back again to the environment, to sensory structures, and so on, during the processing/accomplishment of goal-directed behaviour. Robot perception is flexibly integrated with its own actions, both oculomotor and arm-related. The stereo vision platform integrated in the final system could be potentially substituted by the anthropomorphic mechatronic system that emulates eye kinematics and actuation of the human eye. Though, further functional testing should be necessary in that case.

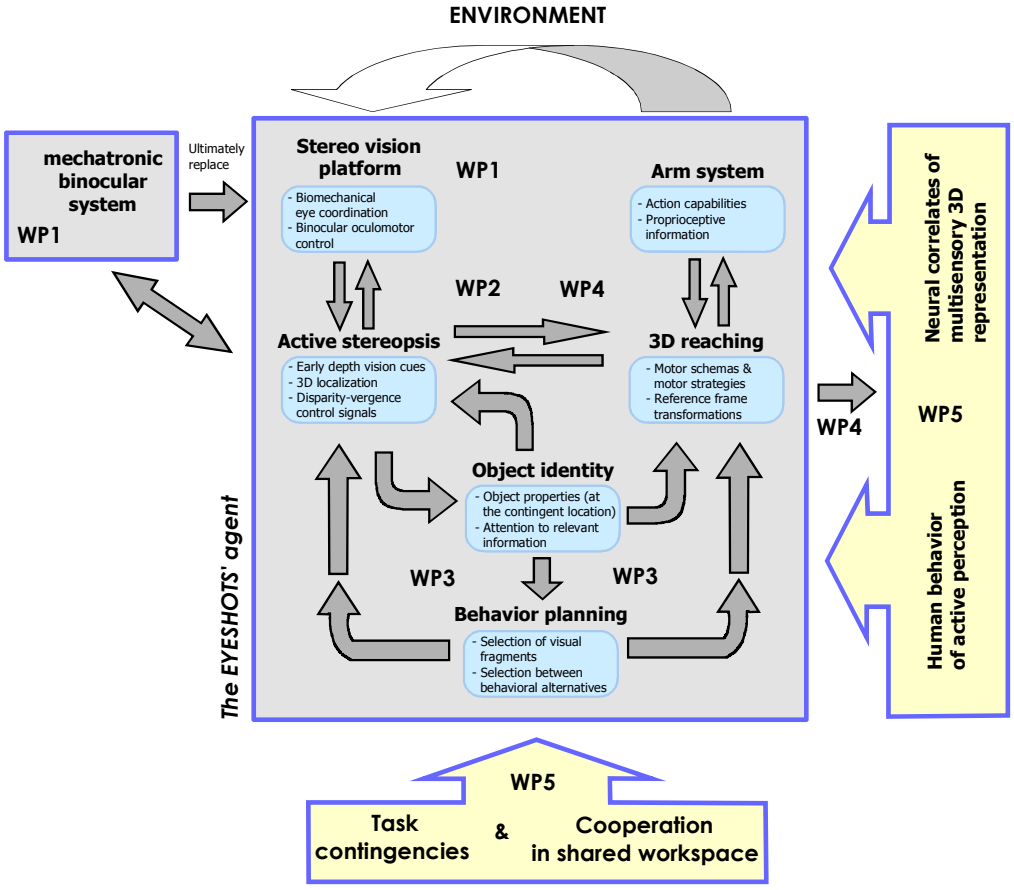


Figure 3: *The EYESHOTS project components and their integration.*

In summary, the project achieved all the objectives of the period 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 **3rd 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* – Analyse the relationships between the version and vergence angles and the resulting disparity patterns. Introduce specific mechanisms to modulate the responses of the disparity detectors on the basis of the vergence and version signals. These modulations are expected to influence perception at a more global level, being integrated across the whole image. The results will be taken into consideration in the design flow of WP2.
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. Experimental validation of the models investigated in Task 1.3. Development of an integrated head-eye-arm platform.

WP2: Active stereopsis

(1) To learn a vergence motor strategy that optimizes the quality and efficiency of the feature-extraction for the specific task to be accomplished.
(2) 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 test the vergence networks developed in the previous prediods in the robot eye system and assess its performance. The networks are trained off-line, using a set of binocular images acquired from the real scene.
Task 2.2: *Interactive depth perception* – To develop a mechanism that renders the disparity-to-depth transformation robust to small vergence/version errors due to the limited accuracy of the motor system. The full disparity vector (i.e., both horizontal and vertical components) will be exploited to obtain accurate eye position information.

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 3.2: *Selecting visual fragment* – Development of dynamic goal-directed attentional selection to bind object properties, on the basis of the momentarily existing task. The final model should be able to localize a particular target object in a visual scene.
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. A model of working memory that allows us to activate context information for the task at hand based on the association of previous events will be developed and tested.

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* – Generate an integrated representation of the peripersonal workspace in a dynamic way, through the practical interaction of the robotic system with the environment. Visuomotor descriptors that combine visual and proprioceptive information concerning eye and arm movement will be 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 7.3: Journal publications, participation to workshops, conferences, and other forum and events. Managing the mailing-list to disseminate results to interested parties.

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.

Summarizing, for what concerns the main S&T issues, the project's objectives for the 3rd reporting period were:

1. Actual development of integrated demonstrators.
2. Physical realization of the tendon-driven mechatronic binocular eye system.
3. Analysis and model validation of the paradigmatic aspects of concurrent and heterogeneous representation of 3D space evidenced in the neurophysiological and psychophysical experiments.
4. Investigate human-robot interaction through action co-representation tasks in shared workspaces.

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

1. The continuous update of the project website.
2. Dissemination activity through the participation to international conferences and workshops, and publication of the results of the research activity in peer-reviewed journals or as contributions to edited books.

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.

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.

2.2 Follow-up of previous review

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

Recommendation no. 1

> **Complete the non-conventional eye system** (Partner UG). The mechanical system offers very interesting perspectives, e.g. exploring the so-called 'structural consequences' of the design to address what is usually considered a 'control problem'. And there may be an interesting use for such a platform in the study of eye pathologies, e.g. strabism.

The mechanical design of the bio-inspired robot-eye has been completed. An integrated version has been implemented and experimental preliminary functional and performance tests have been completed.
The system features the principal theoretic constraints leading to the *hardware* implementation of Listing's Law, furthermore it has been designed to allow for extended experiments based on recently developed control models as part of WP1, (Cannata & Trabucco, 2011 [C17]).
In order to speed up the development time and making possible the conclusion of the integration of the prototype in due time, a rapid prototyping technology has been adopted for the construction of the most critical (complex and expensive) parts, to speed-up construction time, reduce costs, and simplify design updates and variants. The drawings and CAD models will be made publicly available through the project website after the system design material will have been accepted for publication.

Recommendation no. 2

> **Integration of the existing modules (especially the WP1-WP2-WP3 complex) in the UJI demonstrator:** Partner UJI, which is the only partner with access to a complete system (head/eye, torso, arm, hand), needs to develop a demonstrator integrating the novel perceptual and visuomotor processes developed by Partners UG, K.U.Leuven and WWU. Given the short time frame remaining in the project, Partner UJI is to submit a roadmap document outlining how this recommendation is to be implemented. This roadmap should outline a list of tasks by which the different algorithms from the workpackages will be integrated one by one in an order that minimises delays. In sum, the roadmap should include:
>
> I. A list of tasks with a timeline prioritising in which order the different results from WP1 – WP3 are to be integrated into a coherent hard/software system so that delays are minimised. (1 page)
>
> II. A description of the nature of the planned demonstrator, including a scenario. No prescription is made regarding the nature of this scenario but a requirement is that it demonstrates that the developed algorithms work well with real world images and that they are actually running in a concerted manner on a hardware platform. (1 page)
>
> Recommendation 2) is particularly important. Please communicate to the EC Project Officer at latest on 5th July 2010 the requested roadmap.
>
> The reviewers believe that the actual development of an integrated demonstrator would enhance the impact of the project: a successful implementation of the processes developed in a VR environment would demonstrate the validity and scalability of the proposed method to real-world therefore enhancing its impact in a community that is simulation-wary as well as possibly making it an attractive proposition to the industry. Negative results, by which we mean that use in the real world of the developed techniques makes it possible to identify particular scenarios or conditions under which the proposed methods are not as effective as in VR, would open new lines of scientific investigation for partners UG, K.U.Leuven and WWU therefore furthering the scientific impact of this project.

According to the request, on July 5th 2010 the Coordinator submitted to the Project Officer the "*Roadmap for the integration of WP1, WP2 and WP3 modules and algorithms on the UJI humanoid robot setup*" compiled by partner UJI. Before submission, the Roadmap was discussed with the different groups that developed the relevant software components, as well as with the Coordinator and endorsed by the Consortium.

The Roadmap comprises five different integration phases (0-4) specified with expected timings, to which there correspond as many experimental set-ups.

For the sake of clarity the integral version of the Roadmap is reported in Box 1.

Box 1

Castellón de la Plana, July 05, 2010

E. Chinellato, M. Antonelli, B. J. Grzyb, A. P. del Pobil
Robotic Intelligence Laboratory
Universitat Jaume I, Castellón de la Plana, Spain

Roadmap for the integration of WP1, WP2 and WP3 modules and algorithms on the UJI humanoid robot setup

This document describes the integration between the modules of WPs 1, 2 and 3 of project Eyeshots, and the robotic system implemented for WP4. Reference point of the integration process is the humanoid robot of UJI Intelligence Robotic Lab and the behavior implemented on it by UJI partner. According to the workplan in Annex I, at month 30 the model developed by UJI using data and insights from partners UNIBO and WWU (WP5) will be fully functional on the robot platform. Employing the Radial Basis Function framework developed in the model, the robot will be able to achieve open loop gazing and reaching capabilities toward visual targets, according to the goals of Task 4.2. Contextual coding of a target in different reference frames (visual, oculomotor, arm-motor) will allow the system to perform also peripheral reaching actions (without foveating the target). As for the model, the robotic implementation should also emulate some psychophysical effect related to deceptive feedback, such as in the saccadic adaptation paradigm. Modules from WPs 1, 2 and 3 are expected to improve and complement such abilities, and the robotic implementation should serve as a further validation for the modeled functionalities and neural and physiological mechanisms.
The integration process described below is organized in a number of subsequent steps, starting from the modules belonging to the Vergence Version Control model with Attention effects (VVCA), developed by partners UG, K.U.Leuven and WWU/Chemnitz, which are either already available or in debugging phase, and possibly continuing with modules that are at the moment still under development.
We are currently performing, with the aid of partner UG, a preliminary study for setting up utilities for interfacing the different additional modules, in order to solve the problem of compatibility between Simulink (VVCA model) and C++ (robot) platforms. The possibility of having to fully recode some of the modules might delay the development of the plan as scheduled below, and could be subject to the availability of extra human resources by UJI. The reviewers understanding about UJI having already committed to adding such resources, probably based upon the participation to the review meeting of a PhD student who has not been hired using Eyeshots' funds, is not exactly correct. Although we expressed at the meeting our willingness to use external human resources, we cannot guarantee them, since they are bound to other projects with different and specific goals and review procedures.

Phase 0, to be achieved by August, 31, 2010 (month 30), 4PMs
During this phase, the model for building an integrated sensorimotor knowledge of the environment developed in Task 4.2 will be applied on the UJI humanoid robot setup. The robotic implementation of such model includes: 1) a module for visual acquisition and visual processing that generates disparity information regarding simple visual stimuli, such as point-like features on plain backgrounds; 2) the RBF structure described in Deliverable 4.2b for the contextual representation of stimuli in multiple reference frames; 3) modules for controlling the execution of saccadic eye movements by the robot stereo head and arm reaching movements. At the end of this phase, basic functionalities such as **concurrent or decoupled gazing and reaching movements toward simple visual stimuli** will be available to the robot. Each of the following steps will build on this framework to obtain an integrated system with more advanced visual and visuomotor capabilities.

Phase 1, to be achieved by September, 30, 2010 (month 31), 2PMs
In this first integration phase, partner UG's method —for computing binocular disparity inspired on the functionality of area V1— will be adapted to be employed by the robot, and its output will be used by the saccadic control developed in Phase 0. The integration of this module will allow the system to **operate in more complex visual environments, with relatively complex 3D objects and patterned backgrounds**.

Phase 2, to be achieved by October, 31, 2010 (month 32), 2PMs
The open-loop saccadic control of gaze shifts in 3D obtained in phase 0 will be improved in this stage with the addition of a closed-loop vergence control, that will allow for **finer gazing movements upon object surfaces** after the initial saccadic movement has been performed. Two alternative modules are available to implement such functionality, the Convolutional Vergence Control of partner K.U.Leuven and the Dual-Mode Vergence Control of partner UG. Although the first method requires an extensive learning phase, this can be executed off-line on a simulated environment, and the resulting parameters can be transferred to the robot system. The consortium thus agrees in starting the integration with the inclusion of K.U.Leuven's Convolutional vergence control, whilst not excluding the possibility of implementing on the robot both methods for a further comparison between them.

Phase 3, to be achieved by November, 30, 2010 (month 33), 2PMs
This phase is aimed at endowing the system with higher level cognitive abilities. A bio-inspired module for object recognition developed by partner WWU/Chemnitz which takes in input the responses of the V1 module of Phase 1 will be integrated in the framework, together with the Frontal Eye Field (FEF) movement map it provides in output in order to direct the system attention on the selected visual feature. The inclusion of such modules will allow the robot to **work with multi-object setups**, and the system could be required to contextually foveate and reach toward different visual targets.

Box 1 (cont'd)

Phase 4, to be achieved by December, 31, 2010 (month 34), 2PMs
The version control technique developed by partner KUL employs the FEF movement map described in Phase 3 to achieve a precise closed loop control of eye movements around the visual environment. With the inclusion of this module, which will substitute the open loop control of Phase 0, the **full integration of the Vergence Version Control model with Attention effects (VVCA)** onto the robotic platform will be achieved.

Demonstrators and experimental setups

The advancement of the integration process according to the above roadmap can be monitored through the realization of a sequence of tasks in similar but slightly changing and increasingly demanding scenarios.

Phase 0
As a first experimental task, the system can be required to show its visuomotor capabilities by performing an oculomotor action toward a simple, even point-like, target placed in its visual environment, or toward the location where its hand lies (its identification would also be simplified, e.g. with markers). Complementarily, it should also be able to perform arm reaching movements to a similar visual target, either with or without gazing at it. The latter is a case of peripheral reaching, in which an intermediate transformation from visual to oculomotor space is performed, but the corresponding motor signal is not released.

Phase 1
In this phase, more demanding visual targets, such as real 3D objects, can be shown to the system, which has to be able to perform with them the same tasks as in Phase 0. In general, restrictions to the visual conditions should be relaxed in this phase (e.g. requirements on the background, lightning, and similar).

Phase 2
The introduction of fine vergence control requires a further complication of the visual task, in which the system has not only to be able to gaze at objects in the visual space, but also to perform precise, closed-loop vergence movements to focus more exactly on their visible surface.

Phase 3
The next 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 Phase 0 to any visual target, for example in a given sequence or on demand, searching for the goal object by using the Visual Attention module.

Phase 4
The final experimental setup will include all of the above and should allow to enrich the general visuomotor abilities of the humanoid robot, through an improved control of eye movements and advanced visual skills. Ideally, this integrated configuration should be in charge of performing the final experiments on human-robot interaction tasks, to fulfill the last and most comprehensive goal of the Eyeshots project.

Here we summarize (1) *how* the Roadmap has been followed and implemented (progress with respect to the planned timings and final status), as well as (2) the results achieved or the tests performed. A more detailed description of the implementation steps of the Roadmap is reported in Section 3.1.

The development of Phase 0 of the above Roadmap, describing the fundamental skills of the robotic system as implemented by UJI according to the objectives of WP4, has been presented in deliverable D4.3b and will be further detailed in Section 3.2 of this report.
Phases 1-3 of the integration plan have been also fully achieved, as computational modules from the VVCA model of WP1, WP2 and WP3 have been integrated in the UJI sensorimotor robotic framework. The corresponding experimental tests have all been executed following the guidelines of the Roadmap. Full integration has nevertheless been achieved only at the end of the project, and no extra time could be dedicated to the development of Phase 4, which revealed to be especially problematic. Still, the use of the version signal provided in terms of final position calculated by UJI neural networks allowed us to perform in a fully functional manner all behavioral goals of EYESHOTS and of the Roadmap.
Regarding software development issues, for efficiency and homogeneity issues we favored a software integration solution in which all modules are implemented in C++ and installed directly on the robot PC.
Despite the additional initial effort required for porting the modules into the robot platform, this choice revealed to be very appropriate for the final integration, as the external modules were finally integrated with

relative ease into the Yarp robot architecture. In this way, we are now able to recruit either the external or the UJI modules to test each possible configuration of the system on a suitable setup.
Another important issue in the integration of all modules has been the large interocular distance of the robot eyes, which can generate very large disparities that are difficult to manage by the computational modules of the VVCA architecture. This problem has been solved by a careful control of the robot behavior that allows us to maintain disparity within a tractable range, at least for the central image region.
Summarizing, the experiments of the final setup, as described in deliverable D4.3b, will be executed on a relatively complex visual environment with real objects, using biologically-inspired computational solutions deriving from the theoretical studies of partners in WP1, WP2 and WP3. The achievement of the desired robot's behavior close to the project's end prevented us from the use of its full abilities in the human-robot experiments. Hence, results on human-robot interactions have been obtained in a different setting and within a simplified interaction paradigm, as described in Section 3.1 (p. 31) and in Section 3.2 (p.87).

----- o -----

In general, the recommendations have been followed in order to demonstrate in real-world conditions the validity and the scalability of the solutions proposed in the different WPs. That spirit has marked most of the activities conducted in the third period, as detailed in the following section.

3. Work progress and achievements during the period

The work conducted in the 3rd period had a strong technological and engineering emphasis and concentrated on the integration of the different modules on the robot platforms, in order to validate the approach in real-world conditions. Together with these integration activities, the ongoing development of models, as well the analysis of the data collected in neurophysiological and psychophysical experiments yielded significant results on the neuronal mechanisms used to link different fragments by the use of visual, attentional, oculomotor, and arm-movement related cues.
In the following, we will provide (see Parts 1,2,3,4) an overall system framework for the integration and validation activities, specifically focussing on the steps specified in the Roadmap for embedding the novel perceptual and visuomotor processes developed by Partners UG, K.U.Leuven and WWU in the complete UJI humanoid robot setup. Highlights on the major scientific achievements will be presented as well (see Part 5), though referring to the specific Workpackage sections (see the page numbers indicated in the text) for a more detailed description.

3.1 Progress overview and contribution to the research field

Part1: Working modules for stereopsis and oculomotor control across visual fragments
A great effort has been devoted by partners K.U.Leuven, WWU and UG to build an integrated model for Vergence-Version Control with Attention effect (VVCA). The purpose of the VVCA model is to simulate vergence and version control in the presence of an attention signal. As shown in Fig.4, the model consists of:

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

The work, started at the end of the 2nd period, has been completed in the third year by finalizing a set of working modules to be integrated in the eye/head-arm robot system of partner UJI. These modules have enriched, as they became available, the sensorimotor representation of the 3D reaching space developed in WP4. In the following, a descriptive overview of the functionalities of these modules is provided, in the form they have been optimized/designed for the integration. For a detailed description of the VVCA model please refer to Deliverable D2.1(update) and to Section 3.2-WP2.

- *Distributed disparity energy model* [**V1**] – UG.

  The distributed architecture is characterized by two layers of binocular cells (simple and complex cells). The spatial receptive field of each population unit is described by a two-dimensional (2D) Gabor filter. The processing functionality of the simple cells' layer is implemented through separable convolutions. Twenty-four one-dimensional (1D) convolutions for the left and the right images are necessary to build a set of 8 filters to uniformly sample the orientation space from 0 to π.
  From a computational point of view, the use of separable convolution allows us to obtain an improvement of a factor K/3 with respect to the use of the 2D convolutions, where K×K is the spatial support of the filter.
  It is worthy to note that, to avoid the introduction of a loss of balance between the convolutions with the even and odd Gabor filters, the contribution of the DC component is removed. The sensitivity to binocular disparity is then obtained by considering the phase-shift model, that has been implemented trough simple algebraic operations (multiplications and sums) instead of filtering the images with the shifted filters.

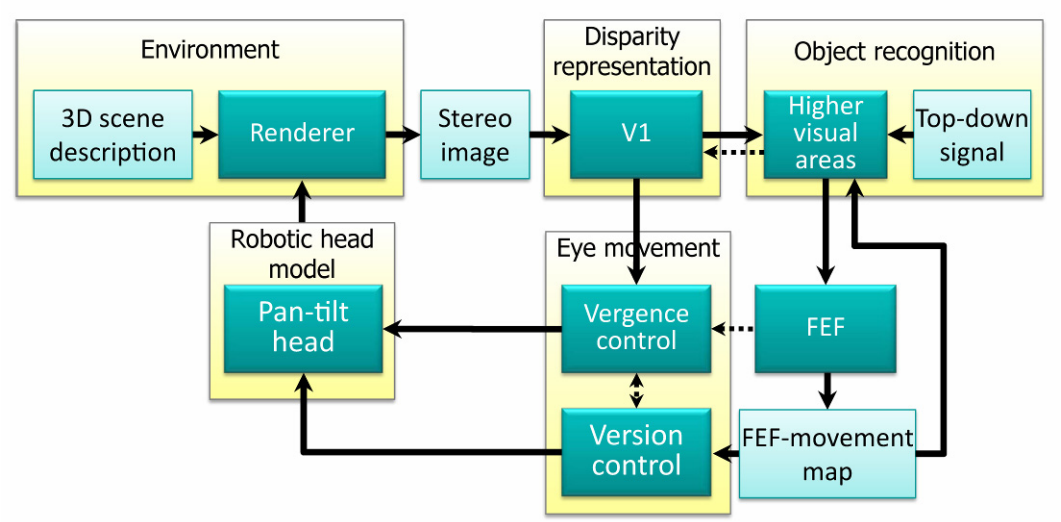


Figure 4: *The block-diagram of the proposed VVCA model. The stereo image generated by the simulator is processed by the disparity detector population, to produce the population response. Depending on which vergence control network is used, the population response is then directed to either the population response post-processing block, which is producing the post-processed population response (the linear VC-net case), or directly to the vergence control network module (the convolutional VC-net case). The (raw/post-processed) population response, together with the actual values of the gaze direction and the vergence angle, are fed into the vergence control network module, the main module of the model. The goal of the VC-net is to produce a new vergence angle, to get the fixation point onto the surface of the object of interest, without changing the gaze direction.*

This approach further improves the performances by a factor of (2×K×P)/(P+2×K) , where P is the number of phase shifts. The performances of the developed solution, as a whole, improve by a factor 16, when considering typical values K=11 and P=7, and by a factor 70 for K=41 and P=7. The response of each complex cell (binocular energy) is then obtained by combining the outputs of the simple cells' layer, through sums and squaring operations.
From a stereo image pair in input, the architecture can be used both for the computation of the binocular disparity and for the exploitation of the distributed representation without an explicit decoding of the cells' responses.
To decode the distributed population response an estimation of the disparity along each of the orientation channels is obtained by a weighted sum, then the full disparity (i.e., horizontal and the vertical components) are obtained from the combination of the information extracted for every spatial orientation channel, by solving the stereo-aperture problem.
The range of the disparity that the model is able to compute depends on the spatial support of the filters. A coarse-to-fine approach can be used in order to increase the range of disparity values.
The model has been developed in MATLAB, then a fast implementation in C++, by using the Intel IPP libraries, has been released.
Given a stereo image pair in input 640×480 pixels, the performances of the algorithm (with 8 oriented filters and 7 phase shifts for each filter), running on an Intel Core i7 2.8 GHz, are:

- 1.4 s to compute the twodimensional disparity map with filters 11×11 pixel, bandwidth 0.833 and radial peak frequency 1/4, 6 spatial scales;
- 0.9 s to compute the twodimensional disparity map with filters 41×41 pixel, bandwidth 0.208 and radial peak frequency 1/16, 1 spatial scales;
- 0.5 s to compute the responses of the complex cells with filters 41×41 pixel, bandwidth 0.208 and radial peak frequency 1/16.

The quantitative evaluation of the algorithm is performed by comparing the results with standard test sequences, for which the ground truth data are available (Scharstein & Szeliski, 2002). It is worth noting that these test beds contain horizontal disparities, only. To benchmark the 2D disparities we have used the dataset described in (Chessa *et al.,* 2011 [J1]).

- *Binocular object detection (and FEF saliency map)* [**ORS**] – WWU.

  The developed module of object detection has been first tested with the VVCA system and then ported to the robot system of partner UJI. The object detection system (see D3.2 or Beuth *et al*., 2010 [C12]) uses learned object representations based on the output of a binoccular Gabor energy model as used in other works. Invariance against disparity and distance while being at the same time selective for different objects is achieved by learning the connections from the energy model to cells of a higher visual order area (HVA), which can be compared to the visual area V4. The learning rule is based on previous work (Wiltschut & Hamker, 2009) and in addition uses a trace rule for learning view-invariant representations of objects (like in Földiak, 1991). 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 HVA cells for finding a particular object in a visual scene. 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 that ensures a preferred processing of the target object already before fixation.

  The model fully achieves the goal of stereoscopic object detection at the robot system of partner UJI (details in section WP3, Task 3.2). Together with the developed model of basal ganglia, the system could also discriminate objects that are determined by their relevance or function.

  The object recognition module has two interfaces: it receives input from the energy model and produces a saccade target as an output.

  The model combines several basic ideas (stereoscopy, biologic plausibility, object recognition, visual search in cluttered scenes) together. Each of them can be found in several models, but so far no model has combined all of them in a unifying approach. Saliency-based models typically do not integrate task relevance and thus they require an exhaustive search in the scene to detect the target object. Concerning stereoscopic object recognition, van Dijck (1999) achieves stereoscopic recognition by first unifying the views edge based and then matches the result against a 3D model of an object. In addition to the lack of biological plausibility, the early unification results in a high number of false responses, which probably decrease the robustness of the system. In comparison to monocular biologically motivated models of object recognition, other systems do not implement learning (e.g. HMAX from Serre *et al.*, 2007 or Hamker, 2005) or they do not robustly recognize scenes cluttered with several objects (e.g. VisNet from Rolls and Stringer, 2001 or LeNet from LeCun, 1998).

- *Vergence control modules*

  - Specializing dual-mode servos [**Dual-mode**] – UG.

    The Dual-Mode vergence control module implements a binocular coordination of the eyes in order to align the optical axes on the surface of the fixed object. The input parameters are: the baseline, the field of view, the focal length of the cameras, and resolution of the CCD. The module provides separately for each eye a vector control $\omega_{L/R}$ for both the horizontal and the vertical alignment of the eyes, on the basis of the average population response to the stereo image flow in the foveal part of the image only (a circle of ~5° of radius).
    Depending on the implementation, the control can be used as a position control or as a velocity control for the rotation of the eyes. The final angular control signal depends on the disparity-vergence linear servos that results from a weighted combination of the population response, and on the geometrical characteristics of the optical system. The weights are determined off-line to approximate desired disparity-vergence behaviors.
    Since the stimulus disparity is exploited as an error signal that decreases step by step during vergence movements, the module is able to provide an effective vergence control irrespective of the geometry of the real system. In fact, in a simulated environment, the module, yileds a proper control for different robotic head models: a tilt-pan model (common and fixed tilt axis for both eyes), a pan-tilt model (fixed vertical rotation axis), so as more complicated models like a tendon-driven head with spherical eyeballs, that follows the Listing's Law (Biamino *et al*., 2005). More precisely, the module, featuring a bifunctional behaviour (cf. the Dual-Mode theory (Hung *et al*., 1986)), provides a LONG control that produces wide movements for large

disparities, and a SHORT control that produces precise movements and stable fixations for small disparities.
Resorting to a normalization stage of the population response (Fleet *et al*., 1996), the control results stable in any condition of natural illumination and effective regardless of the texture of the objects of interest, *i.e.*, those located in the foveal part of the image.

Range and performance: From an operative point of view, the module is able to trigger the correct vergence movement well beyond the theoretical size-disparity correlation limit, which bounds the correct disparity estimation inside a range defined by the angular size of the receptive fields. For instance, for an image of 120 × 160 pixels, that corresponds to a field of view of 22.5° × 30° (for a focal length of 4mm), if we use receptive fields of 43 × 43 pixels (*i.e.*, ~8° × 8°), the correct disparity estimate is confined within ±1.5°, whereas vergence control is still effective in a range of ±6°. For a wider range of disparities the technique can be extended to include a multi-scale and/or a space-variant approach (see also Section 3.2-WP2).

Software implementation: The module was first developed in MATLAB/SIMULINK to be tested in a simulation environment, thus with no regard of the computational time. With an image of 120 × 160 pixel, and on a Pentium Dual Core 2, 2.41 GHz, the MATLAB/SIMULINK version is able to work at a speed of 1.4 frame per second. The source code was first released to partners on the project's website in February 2009 and then upgraded in March 2010. For real-time control the module was ported in C/C++, resorting to the Intel IPP libraries, in order to maximally exploit the computation capabilities of the CPU. Evaluation testing with images of the same size, and on the same CPU, the C/C++ version of the Dual-Mode module is able to work at a speed of 15 frames per second.

Comparative assessment with the state of the art: The Dual-Mode module proposes a biological control to the problem of vergence, different from computer vision approaches, like the *cepstral filtering* (Taylor *et al*., 1994) or a *hierarchical segmentation* of the image (Marfil *et al*., 2003). Starting from a distributed approach, instead of extracting the disparity map (Theimer & Mallot, 1994; Patel *et al*., 1996), the module derives the vergence control directly from the population response, without an explicit computation of the disparity maps (Gibaldi *et al*., 2010a), thus decreasing the computational load, and extending its range of effectiveness. Exploiting the responses of a set of differently oriented binocular energy cells, the model provides a control for both the horizontal and the vertical alignment of the eyes. Instead of decomposing the vector disparity estimate (Rambold & Miles, 2008) to obtain the horizontal (vertical) drive signals, the vergence controls are obtained concurrently by a different mapping of the same population response that take into account the different statistical ranges of horizontal and vertical disparities experienced by a fixating observer. In this way, the model is able to work in real time with natural textures stimuli, much more complicated than vertical bars (Tsang *et al*., 2008; Wang & Shi, 2010).

- Learning vergence behaviors [**VC-nets**] – K.U.Leuven with UG.

The goal of the vergence control module is to produce a control signal for the eyes to bring and keep the fixation point on the surface of the object of interest without changing the gaze direction. Since the task is to nullify the disparity in fovea, vergence control module has input from the disparity detector population response (V1) and converts it into the speed of each eye rotation around pan axis ($\omega_{LE/h}$, $\omega_{RE/h}$). In this model we adopted symmetrical ($\omega_{LE/h} = -\omega_{RE/h}$) strategy, which makes the vergence control independent of the gaze direction. Due to established (and fixed) interfaces with other modules (V1, Robotic Head Model), the vergence control module can be easily represented by different vergence models, *i.e.*, Dual-Mode (provided by partner UG) or Convolutional network based (provided by partner K.U.Leuven).

For the real-world robotic setup (at UJI) the vergence control has been implemented as a closed-loop position control: VC module has the same input (V1 response) as in the VVCA architecture, but the output is the *vergence error* $\Delta\delta$ (the difference between the desired (δ) and the actual (α) vergence angle). The estimated vergence error is then transformed into *updates* of the left and right eye *pan angles*: $\Delta p_{RE/h} = -\Delta p_{LE/h} = \Delta\delta/2$.

As the accuracy of the vergence control depends on the quality of the V1 response, the VC-net can reliably operate in a certain range of target disparities (for which the V1 responses are adequate). In the VVCA model with the baseline 70 mm, the filters of size 43×43, and image resolution 192×192 (40°×40°) the single scale V1 can provide reliable information about disparities in the range ±2.5°. At this disparity range the VC module can produce relatively accurate control. At larger disparity range (up to ≈ ±8°) the VC can also be used, but in this case vergence might take several iterations. The real robot setup (UJI) has much larger baseline (270 mm instead of 70 mm), larger view angle (~90° instead of 40°), smaller V1 filter size (11×11 instead of 43×43) and higher resolution (320×240 instead of 192×192). This, in turn, significantly increases the disparity range for the objects in peripersonal space. To overcome this problem, we allowed the VC module to use all five scales of V1 population response.

The vergence control module for VVCA simulator has been developed in MATLAB/SIMULINK. For the demonstrator its simplified version of position-based control has been ported into C++. The training of VC-networks has been done in MATLAB using the Neural Networks Toolbox.

Most of the classic 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, 1998), use as input the target disparity, which is defined as the difference between the desired and the actual vergence angles. In the proposed model, we do not use the target disparity, but the foveal images of the eyes as input to the vergence control model. Theimer & Mallot (1994) use a multiscale phase-based approach to compute dense disparity maps. The vergence is adapted in order to minimize the global disparity, albeit that the system fixates at the "average" depth of the scene. Hansen & Sommer (1996) follow a similar multiscale approach to estimate the horizontal disparity map. The median disparity of the central area of the disparity map is then used for an asymmetrical vergence control. Stürzl and colleagues (2002) also compute the full disparity map, using responses of complex (position-shift type) horizontal disparity-tuned neurons, for a symmetrical vergence control. In (Marfil *et al.*, 2003), a hierarchical segmentation of the stereo image is computed prior to the estimation of the disparity map, which is then used for a combined vergence/version control. The object nearest to the head is selected as an object of interest (disparity of which is to be nullified).

Part 2: Validation of interactive stereopsis behaviour in real-world situations

- *Dual-mode vergence control on iCub platform* **[UG]**

  Many are the advantages of working with a simulated environment. First it is possible to have a complete knowledge of the geometry of the environment, without any possible source of error, except for perturbations intentionally introduced by the programmer. In a simulation, the position of the objects in the scene, the position of the cameras and their orientation in space, are specified with an absolute precision. Hence, it is possible to compute the projections of the scene on the left and the right retinas/cameras, and obtain the ground truth disparity maps. Second, we do not need the control algorithm to work in real time, in fact, since the time line is simulated, we can afford any computational load between two successive frames.

  In order to test and validate our approach on a real robot stereo head, it is necessary to face two problems: the inaccuracies of the motor systems and the real-time processing demand. In a real robot head the geometry of the system is not as precise as in simulated one, the motors' backlash prevent the repeatability of an experiment and the same control will move the robot in slightly different positions. Moreover, since the algorithm must be able to control the robot in real time, it is necessary to decrease the computational time for the single frame, to make the system work faster, while keeping the same performance and stability.

  Having this in mind, we evaluated the efficacy of the dual-mode vergence control on the iCub stereo platform, which, being designed to mimic a human head, is an ideal platform for the validation of the algorithm. The iCub head is equipped with two DragonFly cameras (Point-Grey-Research) with a focal length of 6mm, and a resolution of 1024 × 768 pixels, which are able to work at a frame rate up to 30fps. The baseline of 70mm well approximates the average interpupillary distance of humans.

The dual-mode vergence control is computed by the C/C++ version of the algorithm previously tested in the simulated environment.

Experimental setup

Three tasks were considered to validate system performance: (1) reach a stable fixation of the surface of the foreground stimulus, presented along the binocular line of sight, (2) shift the fixation point on the background stimulus when the foreground is removed, and (3) follow the foreground stimulus when it is moving in depth (see Fig 5A). When the robot is fixating correctly, the binocular image is characterized by zero disparity in the fovea and the optical axes intersect on the same point on the surface. Since the foreground stimulus is positioned and moved manually, it is not straightforward to measure the depth of the stimulus. To make this data available, and quantitatively assess the experiment results, we used a Kinect sensor device (http://www.xbox.com/en-US/kinect/), which is endowed with a range camera, developed by PrimeSense that interprets the 3D scene information from a continuously-projected infrared structured light. The device is able to work as a 3D scanner system, and thus to produce a ground truth of the depth of the scene with good precision and at a frame rate up to 30fps.

The *hardware* components that constitutes the system are:

- iCub stereo head
  - 2 DragonFly cameras.
  - Motorola Freescale DSP 56F807.
  - ESD - USB to CAN Interface.
  - 3 Faulhaber DC motors 1319T012SR.
- Kinect sensor device.
- Standard PC with an Intel Core2 CPU 6600 @2.40GHz, and 4Gb of RAM.

The *software* tools integrated to make the system work in real time are:

- Microsoft Visual Studio 2008 Professional Edition.
- Integrated Performance Primitives (Intel IPP): multi-threaded library of functions for multimedia and data processing applications, used for image filtering and elaboration.
- NTCan.net library: library supporting the Windows .NET framework, for the communication with the iCub head motors via CAN.
- OpenCV library: Open Source Computer Vision library of programming functions for real time computer vision, used for image visualization and saving.
- OpenNi library: library of interface for physical devices and for middleware components, used to interface with Kinect device sensor.

Results

In the first experiment, we tested the capability of the system to provide a stable fixation on a steady stimulus, and to shift the fixation point. The foreground stimulus is presented at different depths starts that vary, from trial to trial, from 600mm to 1000mm, while the background is at a fixed depth of 1400 mm[3]. Analyzing a single trial, in particular when the foreground stimulus starts at a depth of 600 mm (see Fig. 5B), it is possible to evaluate the behavior of the model. At time 0 the fixation point is supposed to be on the target foreground object, and at the instant when the object is removed, the fixation point is still at the depth of 600 mm, while the background stimulus, which now is covering the entire field of view, is at a disparity that is far outside the detectability range. This implies that, as expected, the depth of the plane is not perceived correctly, and the estimation of disparity would not be able to guide a correct vergence movement. Besides, the vergence control is able to produce a movement, in a rather short response time, toward vergence angle suitable for perceiving correctly the stimulus, and when the cameras are again steady, the fixation point is at the correct depth.

Both at the start and at the end of the trial the fixation is precise and stable, and the error given by the backlash on the motors does not affect the efficacy of vergence. Indeed, working in a visually-based closed loop, the control stops when the disparity in the fovea is reduced to a value proximal to zero, regardless to the real depth of the object, and to the position of the motors, and thus of the cameras. The results show that, even if respect to the simulated environment the precision is lower, the

[3] Depth is measured from the middle point of the baseline of the iCub head.

vergence control, in case of a step of depth is able to discriminate properly the necessity for small movements of the fixation point, in presence of small disparities, so as to produce wider and fast movements in case of large disparities (see Fig. 5B). Moreover it is able to keep the fixation point on a steady stimulus, no matter what is its depth.

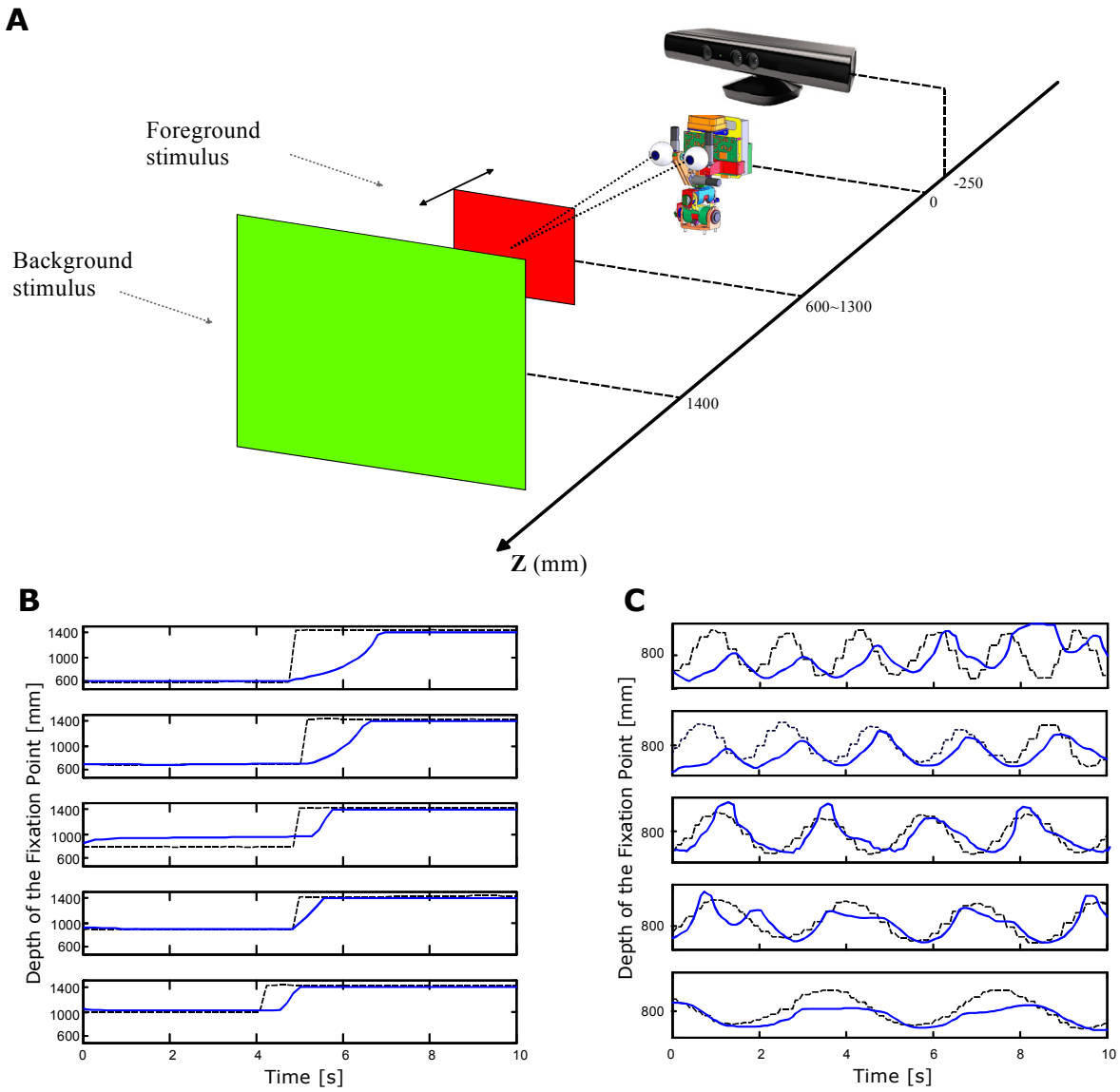


Figure 5: ***(A)*** *The experimental setup used to test the Dual-Mode vergence module on the iCub head. Trajectories of the fixation point (blue solid line) respect to the depth of the stimulus (black dashed line), in case of a step in depth* ***(B)****, and of an oscillating stimulus* ***(C)****. The depth of the stimulus in computed from the middle point of the baseline of the iCub head.*

In the second experiment, we tested the capability of the system to follow in depth a moving object, the foreground target oscillates around a depth of 800 mm, with an amplitude of 200 mm. From trial to trial, the frequency of the oscillation is increased from 30 Hz to 70 Hz. The results are shown in Fig. 5C. When the stimulus is moving slowly (bottom row, f = 30 Hz) the fixation point follows its depth with a small delay, while for a higher speed (top row, f = 70 Hz) the control attempts to achieve the correct movement, but with a minor precision with respect to the previous cases. The trajectories of the fixation point produced by the control for both the step in depth, and for the oscillating stimulus, resemble those that are observed in humans (Hung *et al.* 1986).
To verify that the module is sensitive to the disparity only, the experiments were repeated with foreground stimuli characterized by different textures whose power spectra cover different frequency bandwidths, and in different lighting conditions. Since the model is based on receptive fields tuned

to different orientations, the control is able to cope with more complicated textures than the vertical bars used in (Wang & Shi, 2010), showing an effective texture invariance and an insensitivity to the illumination conditions. Tests conducted for different azimuths and elevations confirmed that the control is able to yield the correct fixation behavior both on a steady and a moving stimulus, regardless of the gaze direction.

- *Disparity and gaze estimation on iCub platform* **[K.U.Leuven and UG]**
  The iCub robotic platform contains a pair of cameras that can pan individually and have a common tilt. The platform used here has a significant tilt offset between the cameras, and we demonstrate here how our autocalibration algorithm can correct this.
  Figure 6 contains three example stereo pairs shown as anaglyphs. The vertical offset is clearly visible and very large 2D disparities occur in the image. We compare the performance of the autocalibration algorithm to a standard two-frame optical flow algorithm (Sabatini *et al*., 2010) that operates on the same multiscale, multi-orientation filterbank responses. For each trial, the autocalibration algorithm was initialized with a (highly erroneous) rectified configuration. The horizontal component of the estimated vector disparity is shown in row B for the autocalibration algorithm, and in row C for the optical flow algorithm. Note that the autocalibration algorithm achieves a much higher density on each occasion. The estimates are also of a much higher precision, as they can be seen by comparing rows D and E. Here, the vertical component of the estimated disparity is shown, and a much more regular pattern is observed in the autocalibration estimates.
  To further demonstrate the correctness of the proposed method, we also show the recovered epipolar geometry in Fig. 7. The red epipolar lines in the right image (B) correspond to the blue keypoints in the left image (A) and vice-versa. It is worth noting that the estimated geometry is very precise everywhere in the image, and also largely different from a rectified configuration.

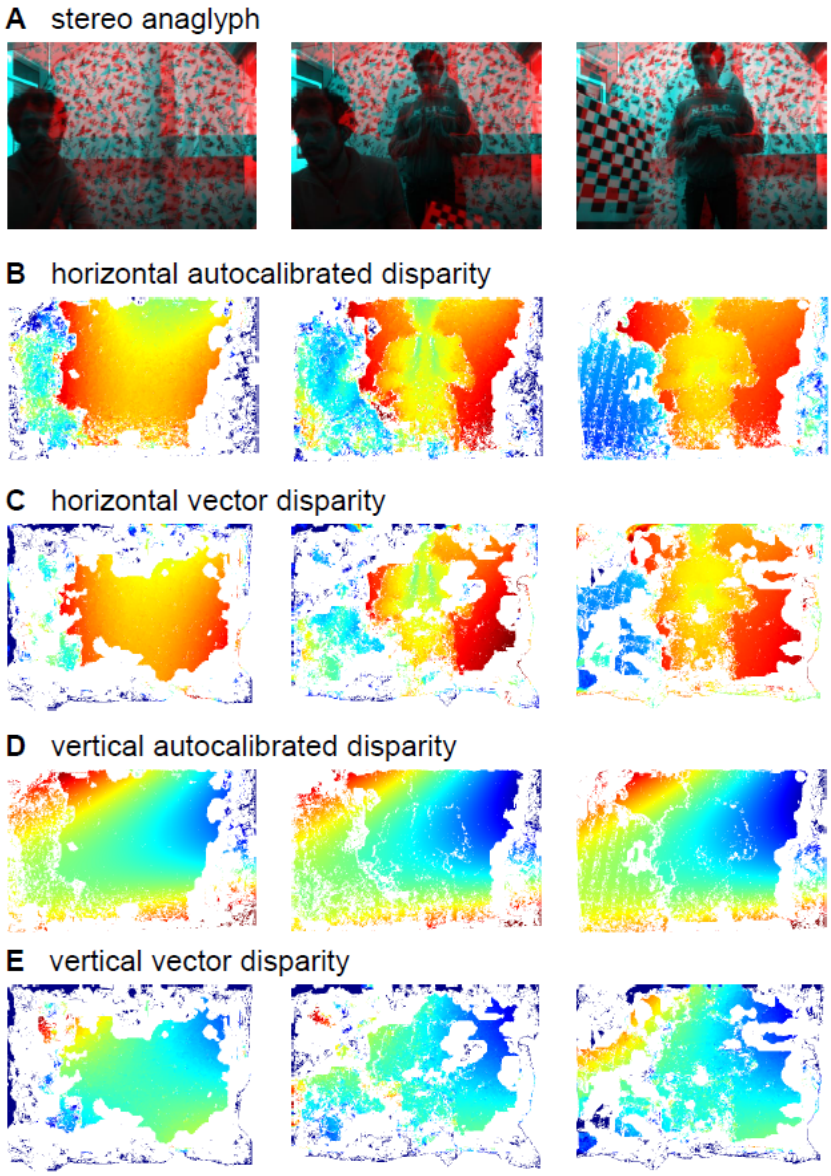


Figure 6: *Disparity estimation results on stereo images obtained with the iCub platform. Rows B and D contain the estimates obtained with autocalibration, and rows C and E contain the estimates obtained with a standard vector disparity algorithm (cf. optical flow).*

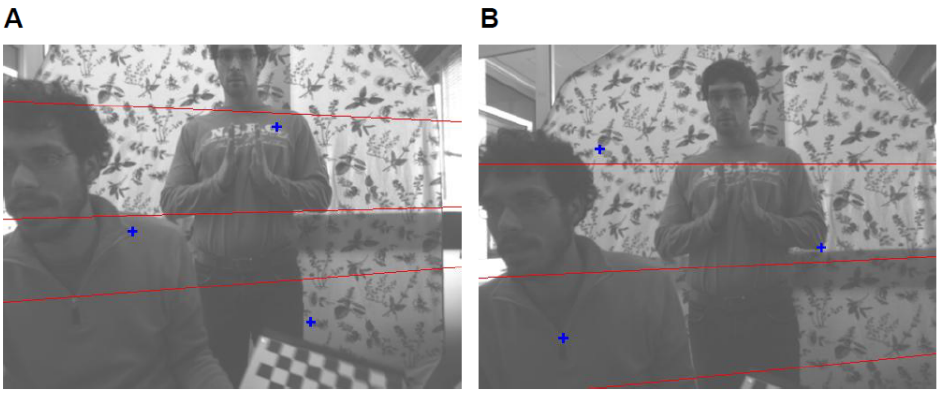


Figure 7: *Recovered epipolar geometry for the scenario of Fig. 4 (center column). Red epipolar lines in the right image (B) correspond to the blue keypoints in the left image (A) and vice-versa.*

Part 3: Realization of the anthropomorphic mechatronic binocular system

The starting point of the project was to complete the understanding of the biomechanics of the ocular motions in humans and primates and to transfer these results into guidelines for the design of robotic eyes which could provide different solutions for the implementation of humanoid robots.

Beside the specific robotic applications it is assumed that the implementation of a bio-inspired robot eye (or robot head) is also the starting point for the analysis and the assessment of the motion control strategies implemented by the brain to drive the very high dynamics of ocular rotation.

In this sense it was, and it is still considered, a key feature of EYESHOTS the target of developing a prototype of robot eye featuring bio-inspired concept and design which are strongly different than the conventional *stiff* pan-tilt platforms. The basic idea is that *emulating* [ocular motions] *is different than simulating* [them]. In other terms is possible with a *conventional* robot system to obtain a desired target behavior by constraining it at control level, it is however, in general, not possible to achieve the same behavior as an emerging one due to the implicit characteristics of the plant.

It is then reasonable to assess, from a pure engineering point of view that state-of-the-art conventional *stiff* robot can guarantee high accuracy and (reasonably) high speed, but they cannot allow to perform experiments where the motion characteristics arise from the intimate nature of the mechanics of the plant.

The rationale adopted throughout the project has been the following.

There exist the evidence of particular types of ocular motions (typically saccades and smooth pursuit) which obey to a basic geometric principle known as Listing's Law. Listing's law specify that the amount of torsion during saccades and smooth pursuit) is zero. This property cannot be achieved by conventional pan-tilt mechanisms unless torsion is properly and actively controlled. However, as the kinematics supporting Listing's Law is not straightforward it emerges the basic question: how Listing compatible motions could arise on a *generic* kinematic structure unless very complex *control circuits* (in a neuro-control framework) or *control models* (in a robot control framework), possibly based on sophisticated sensing, are used?

We have proved that the origin of the characteristics of Listing compatible motions can be grounded on the geometric and mechanical characteristics of the oculo-motor plant. As a matter of fact a reasonably simple model can be defined to achieve Listing's compatible motions independently from the control actions generated by the actuators. This means that by implementing a non conventional robot following the guidelines specified by the models investigated throughout the project is possible to *naturally* achieve ocular behaviours approximating to a large extent the motion of a real eye.

Therefore we have pursued a *strong bio-inspired approach* trying to emulate in the mechanical implementation all the major features arising from the analytical models developed. Despite the simplicity of the model, its implementation on a robot testbed has not been straightforward, and anyway subject to various engineering and technological trade-offs.

The starting point has been that of adopting a tendon driven actuation: ocular motions are generated by distributed forces generated by the actions of the muscolo-tendon system formed by the extra-ocular muscles. We have adopted direct drive brushless motors to emulate the force generators.

The second major step has been that of formulating a complete 3D model of the eye plant compliant with Listings Law. This model allowed to exploit the role of the *soft tissue* surrounding the globe as a visco-elastic element capable of restoring the rest position of the system. An attempt to implement this effect, has been made in the current prototype by including a network of elastic springs.

Finally the last step has been that of miniaturizing the whole system in order to make it appealing for a bio-inspired system fitting into a humanoid platform. The limit here has been the availability of (good quality) miniature commercial cameras and optics. The final eye-ball diameter is in the current prototype is 28 mm (i.e. approx. 4 mm larger than the average human size).

In order to reduce the time and the cost of the final system we have adopted a rapid prototyping manufacturing procedure for the non critical parts (which can be manufactured in about 72 hours of machine time).

Recently published papers have proposed designs trying at different level to use a bio-inspired design. Only the conceptual design proposed in (Mehmood *et al.*, 2008) aims implementing Listings Law, on the basis of our concepts. The others (Villgrattner & Ulbrich, 2011) and (Lenz, 2009) do not have kinematics compatible with Listing's Law. It is anyway worth mentioning the impressive dynamic performances and compact design of the system presented in (Villgrattner & Ulbrich, 2010); this robot features 3dof allowing in principle to simulate Listings' Law. Both (Villgrattner & Ulbrich, 2010) and (Villgrattner & Ulbrich, 2011) feature very high speed piezo-electric motor. Despite their very high cost and complex (and large) control electronics seem the most promising solution for implementing compact tendon driven actuation systems.

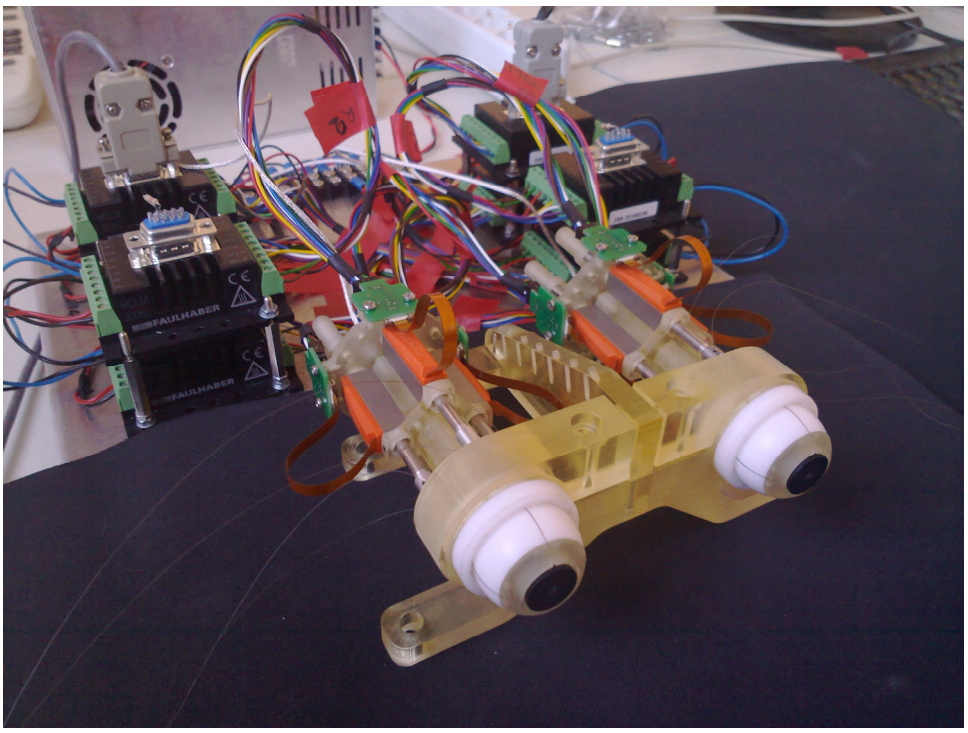


Figure 8: *Picture of the first robot prototype.*

Part 4: Integration of perceptual/visuomotor strategies (WP1-WP2-WP3) on the eye/head-arm UJI robot platform

Basic visuomotor skills of the UJI humanoid robot

The fundamental visuomotor abilities of the UJI robot *Tobatossals* at the end of the project are the subject of deliverable D4.3b. The theoretical aspects of the underlying computational framework and the conceptual development of the robotic implementation have been introduced instead in previous reports and publications. Here, we provide a brief description on the basic implementation schema that allows the robot to act in the experimental setups described in the Roadmap. According to the implementation schema depicted in Fig. 9, there are three sensory/actuation blocks: *cameras* refers to the gathering of visual information by the stereo visual system; *head* represents oculomotor functions, both as eye movements and corresponding proprioceptive information; *arm* deals with reaching movements by controlling the arm joint space.
Transformation blocks *Visual/Oculomotor* (V⇨O) and *Oculomotor/Joint Space* (O⇨A, A⇨O) refer to the radial basis function structures that implement the sensorimotor coordination of the whole system and make it adaptable to the environment and to its own body. The *Visuomotor Memory* maintains the record of previous visuomotor states that allow the robot to code for encountered objects in a way suitable for recognition, searching and for performing gazing and/or reaching actions toward memorized targets. The green blocks represent different visual processing utilities, required to properly interface the *cameras* with the red behavioral modules.

Below, we explain how the computational visual modules of the EYESHOTS' partners have been integrated in the described schema in order to validate them by realizing real-world experiments on a robotic platform, and contextually improve and extend the robot skills.

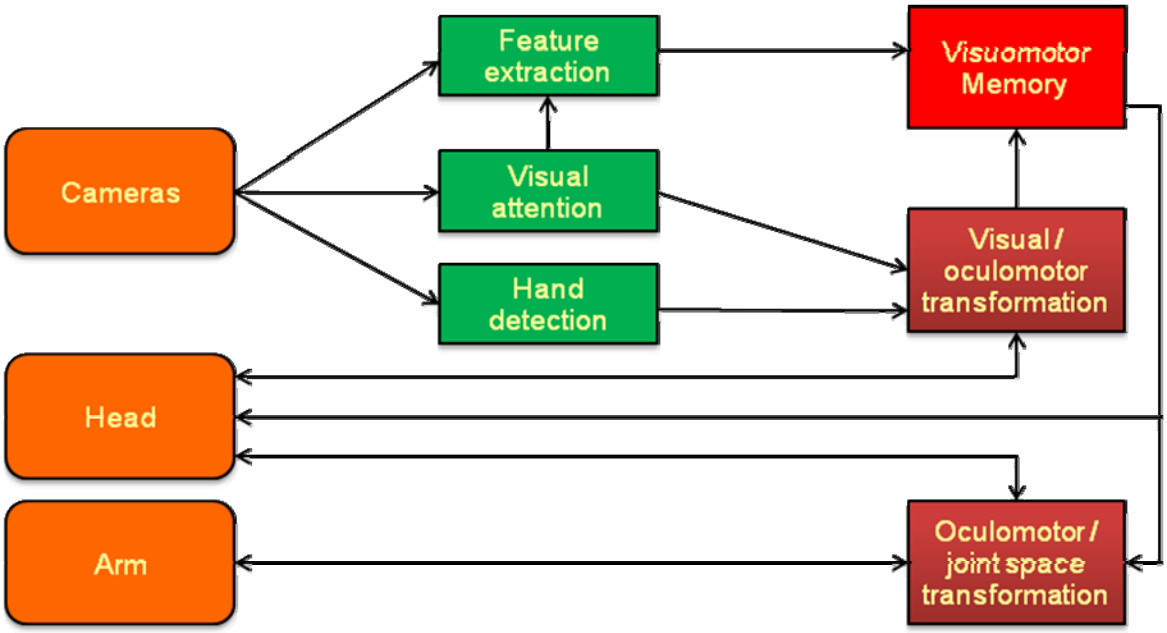


Figure 9: *Implementation schema of the UJI humanoid robot visuomotor skills.*

Integration of the computational modules of partners UG/K.U.Leuven/WWU on the UJI robot.

Three main computational modules developed in WP1, WP2 and WP3 have been integrated in the above framework. They derive from the Vergence-Version Control model with Attention effects (VVCA), developed by partners UG, K.U.Leuven and WWU. The way such modules change the robot visuomotor behavior implementation schema can be observed in Fig. 10. The three modules described in Part 1 of this Section that have been introduced in the schema are **V1**, **VC-nets** and **ORS**.

A first fundmental choice in the integration process was to decide what programming language to use for the modules, as all components of the VVCA architecture are based on MATLAB/SIMULINK, whilst the robot components are all implemented in C++. Although a utility for integration across different platforms had been provided by partner UG, for efficiency and homogeneity issues we finally favored a software integration solution in which all modules are implemented in C++ and installed directly on the robot PC.
The software platform on which all *Tombatossals* software modules are implemented is based on Yarp (Yet another robotic platform, http://eris.liralab.it/yarp/). Yarp provides a set of libraries that help to create the software framework as a collection of stand-alone applications that communicate between them through software ports, e.g. sockets. Each functional block of our framework (V1, VC, ORS, V⇨O, ...) is thus implemented as a program that waits for signals on the input ports and sends data on the output ports. Every module has also a configuration port that can be used to change the parameters or the behavior of the module itself. Two synchronization blocks were included into system to manage the data flow between the functional blocks. The multiplexer block allows us to select which input signals have to be forwarded in output, the task manager block activates the data stream necessary to perform any required task, e.g. saccade to the visual stimulus or look at the reaching point.

Another important issue in the integration of all modules has been the large interocular distance of the robot eyes, which can generate very large disparities that are difficult to manage by the computational modules of the VVCA architecture. This problem has been solved by a careful control of the robot behavior that allows us to maintain disparity within a tractable range, at least for the central image region.
Some details on the integration of each of the three modules are described in the following paragraphs.

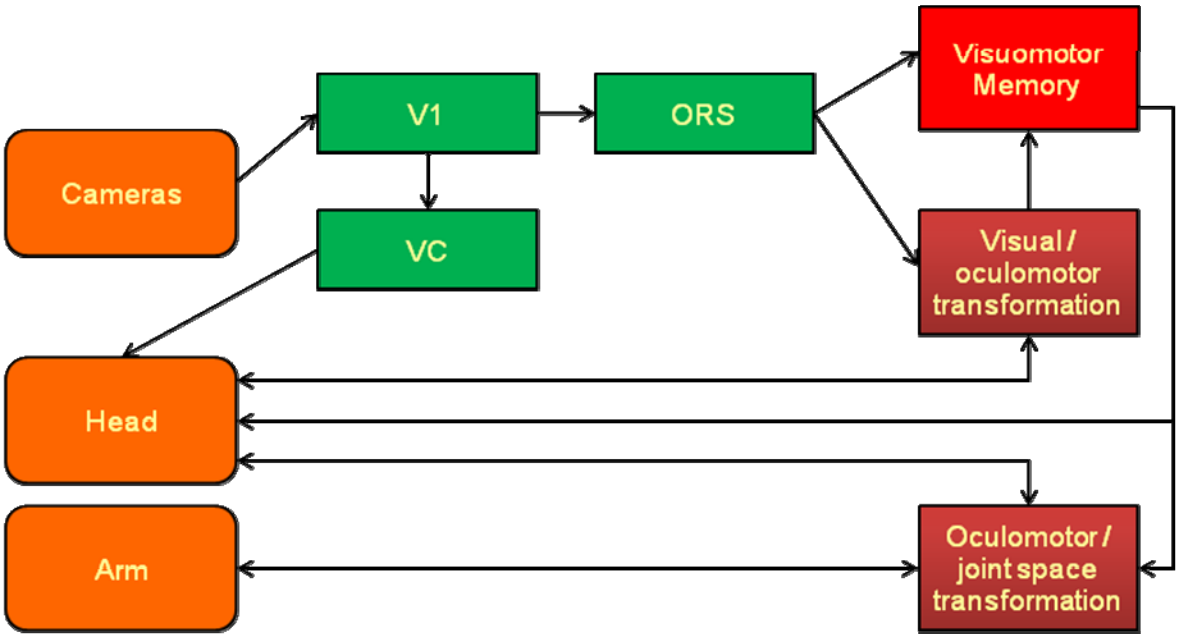


Figure 10: Implementation schema modified by the integration of modules V1, VC and ORS.

Phase 1

In this phase, a C++ version of UG V1 front-end module has been provided by partner UG and integrated by UJI on the robotic system. This module is inspired on the functionality of primate primary visual cortex and computes the binocular energies of a population of Gabor filters that are sensitive to different orientations and interocular phase differences.

Module V1 receives in input the left and right images acquired from the cameras and rescaled at 320x240 pixels. Depending on the overall configuration of the software architecture, the output can be either the population energies or the disparity map. Before the implementation of the next Phases of the Roadmap, the disparity map was used by the UJI visual attention module to detect object location by segmenting the images (textured objects on dark background), and by the UJI open-loop saccadic control to gaze at targets. The centroid of the selected object, together with its disparity, was used as input of the visual to oculomotor transformation. For implementing Phases 2 and 3 the population energies are sent in input to VC and ORS, respectively. In general, the integration of this module allows the system to operate with real 3D objects.

Phase 2

A closed-loop vergence control module is added to the system in this phase. This module has also been implemented in C++ by K.U.Leuven, using a linear vergence control for simplicity. The weights of the network are derived by an off-line training phase, performed by the original MATLAB/SIMULINK version of the module on a training set of real images gathered by the robot. After the execution of a saccadic movement toward a target object, module VC receives in input the current position of the head and the output of V1 in the form of a set of energies extracted by the population of Gabor filters. VC makes use of such visual information in order to perform a finer oculomotor control upon the object surface that brings to zero the disparity in the fixation point. The output of VC is a new target position of the head, aimed at reducing the disparity of the foveated object. In this way, a closed-loop head-camera-V1-VC control is created and remains active until VC output in term of head movement is below a given threshold.

Phase 3

Phase 3 enriches the robot software library with a bio-inspired module for object recognition (ORS) fully developed in C++ by partner WWU. Module ORS receives in input the response of V1 and recognizes the object the system is gazing at, providing the Visuomotor memory with the exact object identity. The module is trained offline with object images taken by the robot from different distances and viewpoints. Module ORS employs the population energies provided by V1 to compute the location of a target stimulus in the left image. For the integration of the ORS in the UJI architecture, an input port for the disparity map has been added to themodule. So, once the target object is detected, its disparity is extracted and binocular information

of the stimulus is sent to the visual to oculomotor transformation for computing a potential saccadic movement toward the target. The selection of the target object to look for is given through the configuration port, and can be either random, sequential or arbitrarily decided by a human user. The robot own hand has been included in the set of objects to recognize, so that the robot is able to identify visually, without the aid of markers, its own limb. The inclusion of the ORS module allows the robot to work with real multi-object setups, such as in the example of Fig. 11.

Phase 4

The integration of the version control implemented by K.U.Leuven in MATLAB/SIMULINK on the robot system was not performed due to technical issues that could not be easily solved. In fact, the version control module should provide a velocity profile to be followed by the robot eyes to focus on the target. Though, the maximum control frequency of the robot head does not allow for the speed required by K.U.Leuven version control module. For this reason, the version signal is provided in terms of final position calculated by UJI neural networks. In any case, despite this last point, all the targeted skills have been attained by the integration of the above three modules on the architecture developed by UJI.

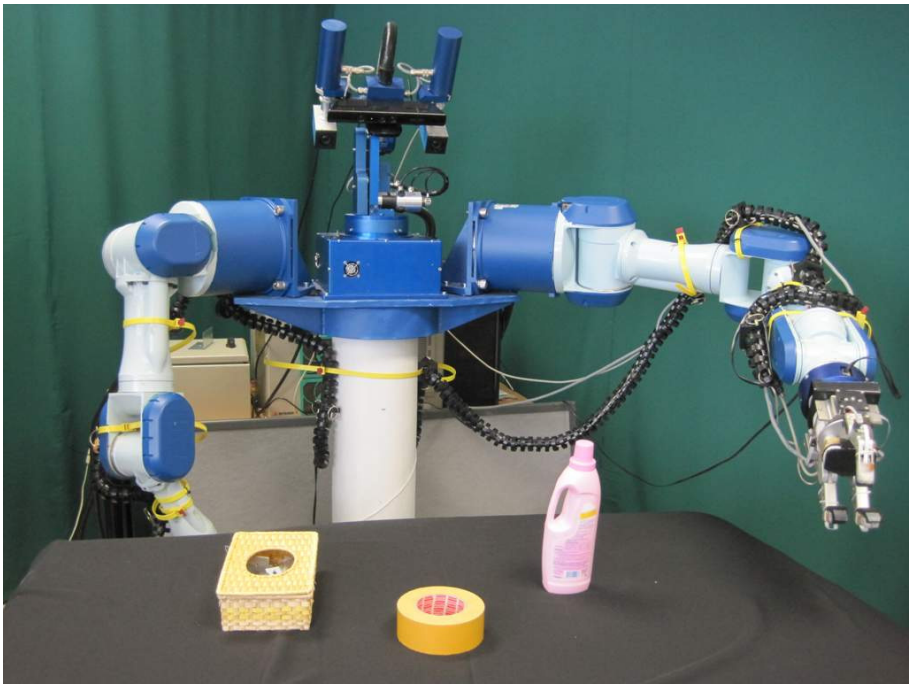


Figure 11: The multi-object experimental setup used with the UJI humanoid robot platform *Tombatossals*.

Summarizing, all but the last of the Roadmap integration phases have been successfully carried on, allowing to enrich the general visuomotor abilities of the humanoid robot, through an improved control of eye movements and advanced visual skills.
On the basis of these results, the experiments of the final setup, as described in deliverable D4.3b, will be executed on a relatively complex visual environment with real objects, using biologically-inspired computational solutions deriving from the theoretical studies of partners in WP1, WP2 and WP3.

Comparing our work with the state of the art, whilst the use of SOM networks in robotics sensorimotor transformations is relatively common (Fuke et al., 2009), the employment of biologically inspired Radial Basis Function (RBF) networks remains relatively unexplored. In fact, to the best of our knowledge, only two papers describe the use of RBF networks for visuomotor transformations. The system of Marjanovic et al. (1996) firstly learns the mapping between image coordinates and the pan/tilt encoder coordinates of the eye motors (the saccade map), and then the mapping of the eye position into arm position (the ballistic map). A similar learning strategy is employed by Sun and Scassellati (2005), which use the difference vector between the target and the hand position in the eye-centered coordinate system without any additional transformational stages. The main difference between these works and ours is that we employ stereo vision, realizing a coordinated control of vergence and version movements. Moreover, the saccade map in

(Marjanovic et al. 1996) is fixed and mainly used to provide visual feedback during the ballistic map learning. On the other hand, our sensorimotor transformations are bidirectional, so that our system learns to gaze towards its hand but also to reach where it is looking at. This skill is trained through a self-supervised learning framework, in which the different modalities supervise each other, and both improve contextually their mapping of the space. The distribution of the RBF centers also differs from the cited works, as we place the neural receptive fields according to findings from neurophysiological studies on monkeys.

A few attempts to tackle the problem of coordinate control of gazing and arm movements by using neural networks, but not RBFs, have also been reported. Schenk et al. (2003) employ a feedforward neural network for learning to saccade toward targets, and a recurrent neural network is employed for executing the transformation carrying from the visual input to an appropriate arm posture, suitable for reaching and grasping a target object. The reaching model of Nori et al. (2007) consists in learning a motor-motor map to direct the hand close to the fixated object, and then activate a closed loop controller that using visual distance between the hand and the target improves reaching accuracy. Eye gazing control is not adaptive, and they do not consider the importance of contextually maintaining a series of representations in different body reference frames, as suggested by neuroscience findings, especially those regarding posterior parietal area V6A.

Finally, the recent work from Hulse et al. (2010), which deals with strategies for coordinating gaze and arm movements, is not directly comparable to ours. The main difference is that they do not use stereo vision, which is at the core of our approach. Moreover, the configuration of their system does not allow for a clear definition of peripersonal space, as their robot is composed by a 2 d.o.f. gazing head and a decoupled 5 d.o.f. arm. Finally, their biological inspiration is developmental and does not include any connectionist modules inspired on neural data.

Part 5: Highlights

- *Statistics of the disparity patterns in the peripersonal space* [UG], see also p. 39.
  In natural viewing conditions, the disparity distributions critically depend on the 3D structure of the scene as well as on the relative orientation of the eyes. Experimental and theoretical work suggest that observers internalize these environmental statistics in the form of a prior for distance and base their distance judgments primarily on this information when faced with decreased or uncertain sensory information (e.g., Chown, 1999). In particular, we expect the relative orientation of the eyes greatly influence binocular disparity patterns for large vergence angles, as they occur during natural visuomotor interaction in the peripersonal space (<1m), whilst the effect is negligible in far viewing condition. Previous results on disparity statistics in natural scenes (Yang & Purves, 2003; Liu et al., 2008; see also Hibbard, 2007) lack of systematic data in the peripersonal space and focus on the disparity distribution over the entire retinal image, rather than on statistical distributions as a function retinal position, and for different gaze directions. By exploiting a high precision 3D laser scanner, we constructed hundreds of registered VRML scenes, by combination of a large number of scanned natural objects, with an accuracy of 0.1mm. Using the available range maps and simulating distributions of binocular fixations through the Active Vision Simulator developed in the first year (Chessa et al., 2009a, 2011 [J1]), we computed the statistics of the disparity patterns for different fixation points and for different eye movements strategies: from the classical Helmholtz and Fick system, to the more biological Listing system and its binocular extension. The study characterizes the disparity patterns that are likely to be experienced by a binocular vergent system engaged in natural viewing in peripersonal workspace, and discusses the implications on possible optimal arrangements of cortical disparity detectors to compensate the predictable disparity components due to epipolar geometry. In particular, redistributing the coverage sensitivity of the cell's population on the basis of the known gaze direction, it is possible to exploit the information coming from the statistics, in order to allocate the resources in an optimal way to obtain a reliable disparity estimate with a minimal number of binocular energy detectors. In addition, the analysis reveals that on average the modulus of the disparity increases linearly with the eccentricity. This is in well accordance with the space variance of the retinocortical log-polar mapping (Lindeberg & Florack, 1994), since the linear increase of the receptive fields size with respect to the eccentricity is necessary to match the linear increase of the disparity.

- *Working memory model* [WWU], see also p. 61.
  We have demonstrated (see deliverable D3.3b) two novel improvements of our Working Memory (WM) model version reported in deliverable D3.3a. First, we have demonstrated the interaction with the object detection system (as described in D3.2). Second, we have developed a fully autonomous model for learning working memory by interacting on a task. One of the goals of the project EYESHOTS is to develop a perceptual agent for sharing a peripersonal workspace. This task requires to hold previously visible information in memory to allow the agent to be able to choose between behavioural alternatives (stimulus-reward-associations). Both requirements are addressed by the proposed WM model. To ensure that a task is general enough and also replicable, we decided to use a well known task from WM literature. In this 1-2-AX task (O'Reilly & Frank, 2006), decisions must be taken dependently on previously presented symbols and the agent must be able to deal with irrelevant objects for the current task. Only special combinations of symbols (e.g. '1', followed by '2' and by 'X') result in one certain decision, all other combinations result in another behavioural alternative. The number of possible combinations is very high and the agent does not know in advance if a symbol is important or irrelevant. This is also typical for real world tasks resulting in a higher difficulty.

  Secondly, we have proposed a biological meaningful foundation of WM. We focus here on the role of the looped architecture of cortex, basal ganglia (BG) and thalamus in controlling WM and motor selection (Haber, 2003; Voorn *et al.*, 2004). Loops including the prefrontal cortex control WM by flexibly switching between maintenance and updating of information. Then, they bias a motor cortex loop to decide between a set of possible responses. The WM model learns to maximize the received reward for a task by the estimates of the expected reward for each symbol. If the model receives more reward than expected, the model reinforces (modulated by Dopamine) connections in a certain PFC loop, which in turn reinforces the memorization of a certain object. The idea is that if the object was helpful to solve the current task, it will also be useful in the future and therefore the model should remember it. Importantly, we have shown that both systems, working memory control and response selection can, develop on the top of the same cortico-BG-thalamic architecture by Hebbian- and Dopamine-based learning.
  A prominent account of the role of BG in WM is the Prefrontal Cortex Basal Ganglia Working Memory (PBWM) model (O'Reilly & Frank, 2006). We see two main differences between their BG model and ours: First, the WM in the PBWM model is prerouted and the BG acts like a gate for fixed memory slots. In contrast, we assume that the whole cortico-BG-thalamic loop maintains WM content and the loops are not fixed to represent a certain symbol. Second, the PBWM model randomly gates stimuli into WM, therefore the system tries out more or less randomly possible strategies of stimulus maintenance to find the correct solution. Our model in contrast is trained by splitting the task in several steps (shaping), which reduces the number of possible strategies. Hence, the PBWM model needs a much higher number of trials to learn the 1-2-AX task (about 30000 compared to 1000 trials in our model).
  Brown et al., (2004) present an account of how cortico-BG-thalamic loops assist in deciding between reactive and planned behaviours. Their TELOS model manages to learn several saccadic tasks and explains single-cell data from various cortical and subcortical brain areas. Compared to other models of cortico-BG-thalamic loops, including ours, their account offers much anatomical detail. Their model contains most known connections between cortex, BG and thalamus and distinguishes between different cortical layers. Yet the WM is modelled as a hard-coded entity that is anatomically restricted to PFC: Visual representations are predetermined to be gated in WM when PFC activity surmounts a certain threshold and to be deleted from it when the next high active input appears. For further comparisons see Schroll *et al.*, (2011) [J23].

- *Modulation of ongoing neuronal activity in the medial parieto-occipital area V6A by covert attention shifts* [UNIBO], see also p. 73.
  Link across single visual fragments can be obtained in many physiological situations. Commonly, in natural conditions, when we catch with vision a target of a potential reaching action, we move the eyes toward it and then the hand. Due to less inertia of the eyes, the eyes land on the target well before the hand starts to move. In area V6A of the medial parieto-occipital cortex, we have found neurons discharging in this interval, that is in the first 500 ms of fixation of a target in the dark. Interestingly, this kind of cells in V6A strongly prefer targets to be fixated in the peripersonal space,

that is in the reachable space (Hadjidimitrakis *et al*., 2010 [C3]). We interpreted this neural behaviour as the neuronal correlate of a calibration between the eye and the arm systems and we proposed in EYESHOTS that the strong preference for reachable targets in early fixation period could reflect the shift of the attentional spotlight for the purpose of highlighting the location of the target of eye and hand movements in reaching an object (see Hadjidimitrakis *et al*., 2010 [C3]).
Actually, overt deployment of attention is seen in 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.
However, the link between attention and goal-directed motor actions 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, similarly 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.
Attention is important for providing the link across single visual fragments, attention is used to select targets in a visual scene for prioritized processing and for preparing appropriately directed actions, as manual reaching or grasping. Our study intended to measure the influence of covert attention toward different parts of the visual world in neurons of area V6A, a visuomotor area of the dorso-medial visual stream involved in coding reaching movements to targets in space (Galletti *et al*., 2003; Fattori *et al*., 2005).
We induced in the monkey covert shifts of attention in absence of any effector movement, neither of the eyes nor of the arm. We performed single cell recording in V6A, while controlling the monkey focus of attention addressing it toward several positions in the workspace. In this way, we could study the influence of spatially directed attention on neurons in area V6A.
It has been found that the neural modulation was present when the covert attention was shifted without any concurrent shift of the direction of gaze (Galletti *et al*., 2010 [J10]). It has been suggested that this attentional modulation is helpful in guiding the hand during reach-to-grasp movements, particularly when the movements are directed towards non-foveated objects. The covert attentional modulations could allow V6A cells to select the goal of reaching during movement preparation, as well as to maintain encoded, and possibly to update, the spatial coordinates of the object to be reached out during movement execution.

- *Social Simon effect* [WWU], see also p. 87.
  In a new development in the final year we have begun to look into human-robot interaction from the perspective of cognitive science tools that allow to measure the quality of the interaction. This is not directly related to the successful completion of a joined task. Rather, it is related to the question of how acceptable and natural a robot can become as an interaction partner. In social interaction between humans, humans seem to co-represent their partner's actions (Sebanz *et al.*, 2003, 2005, 2006). Action co-representation is typically investigated by using spatial compatibility tasks, like the *Simon Task*, in a cooperative setting. In the *Simon task*, participants carry out spatially defined responses to non-spatial stimulus attributes. Responses are usually faster when the stimulus location and the response location correspond (Simon & Rudell, 1967). This effect disappears when a participant responds to only one of the two stimuli, but reappears when another person takes care of the other response (*Social Simon Task*). This *Social Simon Effect* (SSE) has been considered to provide an index for action co-representation (Sebanz *et al*., 2003).
  We used the SSE as a marker to measure the co-representation between a human and a humanoid robot functioning in a human-like way. We aimed to test if the SSE can be used as a benchmark-tool for the perceived humanness of a robotic system. Experiments were conducted with the UJI humanoid robot *Tombatossals*.

When the robot was described as functioning in a human-like manner, e.g. being able to actively decide when to respond on the basis of a neural network, we observed a reliable and robust SSE. However, when the same robot was described as purely deterministic, e.g. being completely controlled by a computer program, the SSE was diminished.
These findings suggest that action co-representation of non-biological agents (e.g. robots) can occur if an agent is perceived as human-like. Higher order cognitive processes seem to affect if we corepresent the actions of other agents be it humans or technical systems. Further, our results suggest that the SSE can be used as a benchmark-tool for the perceived humanness and acceptability of a technical system.

3.2 Workpackage progress

Here we recall the objectives for the tasks of the third 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: Study of the functional implications for depth vision of the different eye movements strategies (Helmholtz, Listing, and L2) and of their impact on possible optimal arrangements of cortical disparity detectors. The analysis was based on the statistics of the binocural disparity patterns obtained by simulated active fixations in real-world peripersonal workspaces acquired by a laser scanner.
Results: We have derived from a computational point of view (1) the identity of Helmholtz torsions (Tl = Tr), postulated by Tweed (Tweed, 1997), for different instances of the visual constraint; and (2) the proportionality relationship, represented by the factor μ, between the rotation of the Listing's planes and the vergence angle. The analysis of the disparity patterns experienced by an active observer in the peripersonal space evidenced only minor statistical varations among the different eye movement strategies.
Status: The work has been completed as planned
Documentation: Deliverable D1.2(update)
Publications: Canessa et al., 2011a (submitted) [J18], Sabatini et al., 2011a (submitted) [J24], Canessa et al., 2011b (submitted) [C16], Sabatini et al., 2011b (submitted) [C15].
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.
- Derivation of an *exact* 3D model describing the dynamics of Listing's Law based motions. The models takes into account the effects of viscoelastic orbital tissue and its role during ocular rotation control.
- Definition of a 2D computational model for the definition of the EOM tension to regulate the eye orientation in 3D.

Status:

- Extended models developed and implemented in simulator (see task Task1.4).
- Complete analysis of mapping EOM static action forces to 3D ocular orientation.
- Simulation tests and numerical validation of the computational algorithm.

Documentation: deliverable D1.3
Publications: Cannata & Trabucco, 2011 (submitted) [C17].
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:

- Final release of a dynamic simulator for studing ocular mechanics and visual processing and control techniques (Deliverable D1.4a). The simulator is a Simulink Toolbox which will be made publicly available through the project website after the related system results will have been accepted for publication in journals/conf.proceedings.
- Basic concept study of the bio-inspired robot eye.
- Extended concept study of the bio-inspired robot eye (accounting for the last year achievements in Task 1.3)
- Selection of main components for implementation of the eye prototype including:
    - Servo motors
    - Servo amplifiers
    - Embedded USB camera.
- Development of a test rig for actuator and control tests in place. Experimental test to assess performance of commercial components required for the implementation of the robot eye.
- Detailed design of the mechanical components.
- Experimental manufacturing test using rapid prototyping to reduce production costs and time.

Status:

- Simulator tested and running (various users of the UG-MACLAB and within the Consortium have used or are using it for tests).
- Preliminary tests for integrated image based closed loop control simulation performed.
- Integration with VR simulator and control software modules.
- Test rig for actuator and control tests in place.
- Experiment for assessing actuation accuracy completed.
- Complete mechanical design completed. Drawings and CAD models will be made publicly available through the project website after the system design material will have been accepted for publication in journals/conf.proceedings.
- Complete bio-inspired tendon driven prototype implemented.
- Preliminary functional tests performed.

Documentation: deliverables D1.4a, and D1.4 (submitted in September 2010, and available also in a revised form, which includes the characterization of the final prototype).

Publications: A manuscript is in preparation (G. Cannata & A. Trabucco) for the IEEE/ASME Transactions on Mechatronics, Focused Section on Bio-Inspired Mechatronics.

Revised planning: 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). Development of the bioinspired robot eye delayed by approximatively one year.

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) Extend the dual-mode vergence control model to log-polar images to extend the range of the vergence control. (2) Test how proper disparity-vergence servos can be directly learned from examples of the desired vergence behavior in real-world conditions. (3) Validate the approach on robotic platforms.

Results: By proper space-variant weighting of the population responses, it has been demonstrated that disparity-vergence responses can be steered to cope with the space-variant epipolar geometry in the transformed cortical domain. The vergence model has been profitably included and tested in the overall VVCA architecture.

Status: The work was completed as planned. The deliverable D2.1(update) was submitted on time.

Documentation: Deliverable D2.1(update). Technical meeting notes by Nikolay Chumerin and Frederik Beuth (Chemnitz, 2-9 January 2010) and by Nikolay Chumerin, Frederik Beuth, Agostino Gibaldi and Andrea Trabucco (Genoa, March 2010).

Publications: Chumerin et al., 2010 [J9], Solari et al., 2011 [J6], Sabatini et al., 2011 (submitted) [J24], Gibaldi et al., 2010 [C2], Gibaldi et al., 2011a [C1], Gibaldi et al., 2011b (submitted) [C14]. A manuscript is in preparation on the integrated model for vergence-version control with attention effect (VVCA) (N. Chumerin, F. Beuth, A. Gibaldi, A. Canessa, M. Van Hulle, S.P. Sabatini & F.H. Hamker).
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 a robot head is not accurate enough, therefore, vision is used to improve upon this.
Scheduling: (month 6-36)
Performed actions: (1) Developed a biologically-inspired algorithm for the transformation of retinal disparity into a 3D scene description based on head-centric disparity; (2) enabled the algorithm to operate directly on the response of a population of binocular energy neurons, (3) applied a learning approach to determine the weights of the neural network that implements the algorithm, (4) applied the same approach to learn gaze estimation directly from the population response, (5) demonstrated the feasibility of the autocalibration procedure on real-world images obtained with the iCub-platform
Status: The work progressed as planned and the deliverable D2.2b was delivered on time.
Documentation: Deliverable D2.2b.
Publications: --
Revised planning: none

WP3: Selecting and binding visual fragments

Task 3.2: Selecting visual fragment
Development of dynamic goal-directed attentional selection to bind object properties.
Scheduling: (month 7-30)
Performed actions: Application and refinement of the developed models to real world scenes taken from the robotic head system.
Results: The performance has been demonstrated on real world scenes. It has been possible to learn slight view invariant representations of objects. In cluttered scenes that contain multiple objects, a selection of the task relevant object has been demonstrated.
Status: Finished as planned.
Documentation: Deliverable D3.2
Publications: Zirnsak *et al*., 2010 [J17], Zirnsak & Hamker, 2010 [J16], Beuth *et al*, 2010 [C12].
Revised planning: none

Task 3.3: Selecting between behavioral alternatives
Learning of the cognitive control of visual perception.
Performed actions: The model of Basal Ganglia has been further refined and connected to the VR simulator.
Results: It has been demonstrated that the model can operate using realistic inputs at the object level and allows us to learn a complex task (1-2-AX).
Status: Finished as planned.
Documentation: Deliverable D3.3b
Publications: Vitay & Hamker, 2010 [J14], Schroll et al., 2011 (submitted) [J23].
Revised planning: none

WP4: Sensorimotor integration

Task 4.2: Generating visuo-motor descriptors of reachable objects
The objective of this task was to implement a model of how to generate an integrated sensorimotor representation of objects in the peripersonal space through the physical 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 oculomotor/arm-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 the learning experiments and define the experimental setup for the real robot.
- Implement the full 3D model on the humanoid robot platform, making the robot able to interact with its peripersonal space through vision, eye and arm movements.

Results:
- The system (simulation and robot) is able to accurately learn the transformation between visual, oculomotor and joint spaces.
- The system (simulation and robot) adapts to altered perception and is able to reproduce some effects of saccadic adaptation.

Status: Completed as planned.
Documentation: Deliverable 4.2a, Deliverable 4.2b.
Publications: Antonelli et al., 2011 [C9], Chinellato et al., 2011a [J2], Chinellato et al. 2011b [J3], Chinellato et al., 2011c [C10].
Revised planning: none

Task 4.3: Constructing a global awareness of the peripersonal space
This task extends the skills of Task 4.2 to the exploration of visual stimuli in the surrounding space. The agent simultaneously learns to reach towards different visual targets, achieving binding capabilities through active exploration, and builds an egocentric "visuomotor map" of the environment.
Scheduling: (month 19-36)
Performed actions:
- Define and implement on *Tombatossals* humanoid robot a global software architecture that allows us to integrate tasks of different level of complexity, either internal or developed by other partners.
- Manage allocation of attention and visuomotor memory in order to perform both gazing and reaching actions and also object recognition by integrating space (dorsal) with identity (ventral) visual information;
- Integrate biologically-inspired modules (on stereo vision, vergence control, object recognition) by other partners in order to validate the underlying theories and provide the robot with enriched visuomotor behavioral capabilities (advanced vergence control, real world objects, ...).
- Perform comprehensive experiments in which the robot is able to operate in multi-object setups creating its own visuomotor awareness of the environment.

Results:
- The robot is able to interact with the objects in its environment, recognizing them and performing coupled or decoupled visuomotor and arm-motor actions;
- The robot is able to create an egocentric representation of its peripersonal space, that allows it to perform custom, goal directed actions toward one of many available targets.

Status: Completed as planned.
Documentation: Deliverable 4.3a, Deliverable 4.3b, Deliverable 4.3c.
Publications: Antonelli et al., 2011 [C9]. A manuscript (M. Antonelli et al.) is in preparation on the full integrated robotic system.
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 2 electrophysiologial experiments.
Results: Both studies are completed. Results have been published as papers of peer reviewed journals (2 in Journal Neuroscience). One additional manuscript is under revision. Results have been proposed to international meetings and shared with the other EYESHOTS' partners.

Status: work concluded as planned.
Documentation: deliverable D5.1(update)
Publications: Bosco *et al.*, 2010 [J8], Galletti *et al.*, 2010 [J10], Gamberini *et al.*, 2011 [J4], Hadjidimitrakis *et al.*, 2011 (submitted) [J19], Hadjidimitrakis *et al.*, 2010 [C3], Gamberini *et al.*, 2010 [C4], Passarelli *et al.,* 2010 [C5], Bosco *et al.*, 2010 [N1], Bosco *et al.*, 2011 [C8]
Revised planning: none

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 data collection ended and the analyses have been performed. Results have been shared with the other EYESHOTS partners and have been used to implement computational models developed in other WPs. One joint paper has been published and another manuscript is in preparation.
Status: work started and conducted as planned.
Documentation: --
Publications: Chinellato *et al.*, 2011a [J2]. A manuscript is in preparation (Breveglieri R, Hadjidimitrakis K, Bosco A, Sabatini S, Galletti C, Fattori P Balanced sampling of visual fragments in the reachable space by parieto-occipital neurons).
Revised planning: none

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: In the third period WWU conducted four behavioural experiments.
Results: Seven studies were completed and are published or in preparation for publication.
Status: Work finished as planned. The cooperative effort between UNIBO and WWU on adaptation experiments on the monkey has been conducted in the third period (Wulff et al., 2011, in preparation).
Documentation: --
Publications: Zimmermann & Lappe, 2011 [J7], Zimmermann & Lappe, 2010 [J15], Havermann *et al.*, 2010 [C6], Schnier et al., 2010 [J13], Schnier & Lappe, 2011 [J21], Galletti et al., 2010 [J10], Havermann et al., 2010 [J11], Havermann et al., 2011 [J20].
Revised planning: none

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 action 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. An additional behavioural experiment on a human-robot interaction was conducted in cooperation between WWU and UJI.
Results: The studies were completed and are currently submitted or in preparation for scientific publications.
Status: Work started and proceeded as planned. Milestone M9.ante (Experimental data in single actor …) was reached as planned on month 18. The human-robot interaction study was conducted on month 36.
Documentation: deliverable D5.4 and D5.4(update)
Publications: Volcic & Lappe (2011) [J22]. Three manuscripts are in preparations (R.Volcic & M. Lappe Gaze behavior in cooperative action, R. Volcic & M. Lappe Predictive eye movements in gaze and action observation, and A. Stenzel, E. Chinellato, M.A. Tirado Bou, A.P. del Pobil, M. Lappe, R. Liepelt How humanoid robots become human-like partners in joint actions).
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 7.3: External dissemination
Scheduling: ongoing
Performed actions: See section 5.2
Results: We have by now published 39 conference contributions and 27 journal papers.

WP8: Training, education and mobility

Task 8.1: Literature database
Scheduling: ongoing update of the database

Task 8.2: Student's seminars
Performed actions: On the basis of the numerous occasions of exchange of knowledge between partners, each student has been asked to write a one-page retrospective of the work done by a cooperation partners.
Documentation: Deliverable 8.2

Task 8.3: Personnel exchange
Scheduling: ongoing
Performed actions: Several visits took place among partners in the reporting period to carry on collaborative research and for the preparation of coauthored manuscripts.

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 (21/19.22) and UJI (2/4.54)
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.

Progress towards objectives

Task 1.2: ***Perceptual influences of non-visual cues***

In natural viewing conditions, the disparity distributions (horizontal and vertical) depend on the orientation of the eyes as well as on the structure of the scene. Previous attempts to analyse the statistical structure of natural scenes and how this structure could influence neural processing and visual percepts are dominated by studies of long-range environments characterized by distances greater than 2m. Yang and Purves (2003) first measured the distribution of distances in 74 scenes (23 fully natural scenes, and 51 scenes containing natural and constructed objects) using a laser range finding technique. Following, other researchers tried to predict the distribution of disparities by extending the model derived by Yang and Purves (cf., Hibbard, 2007) or by using their orginal range data and considering distributions of binocular fixations either measured or simulated (cf., Liu *et al*., 2008). Yet, the binocular disparity has great impact at close distance (<1m), in the so called peripersonal space, considered as the reachable and graspable workspace, for which, to the best of our knowledge, (systematic) data analysis is still lacking.
Moreover, (Hibbard, 2007) and (Liu *et al*., 2008), besides considering viewing distances far from the peripersonal space, focus their attention on the global disparity distribution over the entire retinal image, without considering the gaze direction. Differently, we focused on the disparity distributions that can occur in every retinal location for different gaze directions. From the analysis of the disparity patternsm we considered the implications on possible optimal distributions of the cortical disparity detectors. A priori information on disparity patterns turns out to be very important if we want to model the behaviour of the disparity detectors of primary visual area V1. Indeed, it is possible to exploit the information coming from the statistics, in order to allocate the resources in an optimal way. By redistributing the coverage sensitivity of the cell's population on the basis of the known gaze direction, we expect to improve disparity detection with a reduced amount of resources (i.e., a reduced number of binocular energy units).

Data acquisition – For the simulations shown in the following, we first captured 3D data from a real-world scene by using a 3D laser scanner (Konica Minolta Vivid 910), with the optimal 3D measurement operating range from 0.6m to 1.2m, which is appropriate for analyzing the disparity information experienced by an active observer in his/her peripersonal space. The system allows also capturing the color textures at a resolution of 640 × 480 pixels. Each scan contained up to 307,200 points within a variable field of view, which was adjusted with respect to the size of the object to be scanned. For this work we considered cluttered desks with a collection of hundred real-world objects. The whole scene, as well as the single objects were scanned, registered and merged together to obtain full models of more than 13,000,000 of points each (see Fig. 12A). Off-line registrations of data guarantee an accuracy of about 0.1 mm. A full 360-degree view of the scene is acquired to minimize the occlusion problems that occur when one simulates changes in the vantage point of the virtual observer.

Simulated fixations in the acquired peripersonal scenes – The real-world environment, captured by the 3D laser scanner, is then "explored" through the active vision simulator developed by partner UG in the first

year of the project (Chessa *et al*., 2009a). Such simulator has been implemented in C++, using OpenGL libraries and the Coin3D toolkit (http://www.coin3d.org/) developed for effective 3D graphics rendering. This system is capable of handling the commonly used 3D modeling formats (e.g., VRML), and thus the data acquired by the 3D laser scanner. To obtain the toe-in stereoscopic visualization of the scene, useful to mimic an active stereo vision system rather than to make humans perceive depth, we have modified the SoCamera node of the Coin3D toolkit. Moreover, the developed tool allows us to access the buffers (see Fig.12B) used for the 3D rendering of the scenes. The 3D data and the textures are loaded in the active vision simulator, then the left and right projections, the horizontal and the vertical ground truth disparity maps, are obtained, for each possible fixation point. More details on the simulator are reported in (Chessa *et al.*, 2011 [J1]).
The developed tool has been used to create a database of real-world range data and stereo image pairs for a variety of fixations (see Fig. 12C), in order to guide modeling and for algorithmic and behavioral benchmarks in real-world but fully measured environments. Benchmark images and sequences have been made available to the scientific Community (http://www.pspc.dibe.unige.it/Research/vr.html).

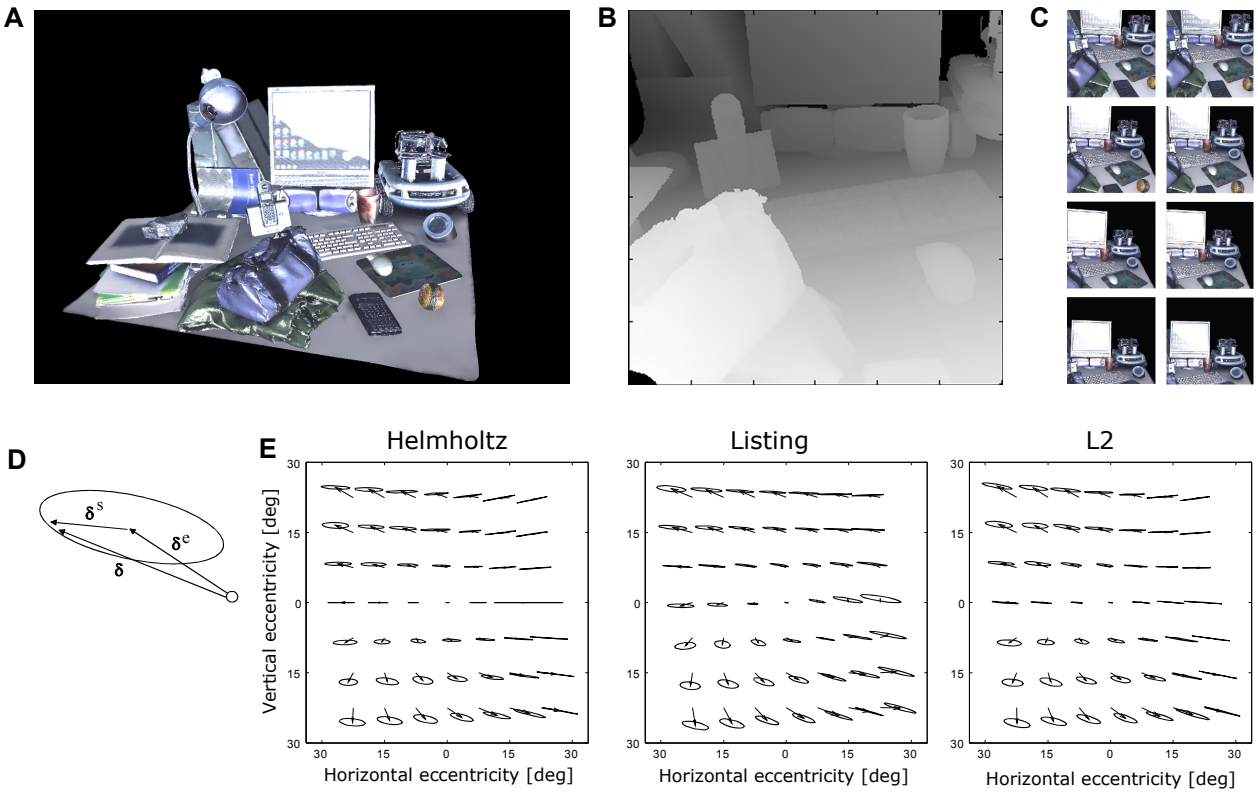


Figure 12: ***(A)*** *An example of real-world scene acquired by the laser scanner. Together with the 3D data the system is able to attach the real color texture to the scanned objects. Two outputs of the active vision system simulator: the Z buffer* ***(B)*** *and the left and right image pairs* ***(C)****.* ***(D)*** *The disparity δ can be divided into two components: one, unpredictable, due to the scene, called residual disparity* δ_s*, and one, predictable, due the geometry of the adopted vision system, called epipolar disparity* δ_e*.* ***(E)*** *The mean vector disparity patterns and the standard ellipses, averaged over all the fixation, for a Helmholtz, a Listing and a L2 system. For the sake of clarity only a grid of 7x7 retinal points is shown.*

Statistical analysis – For a given eye posture we computed the distribution of the horizontal and vertical disparities for all the objects whose images fall within an angle of ±22.5° in both retinas. The other parameters used were: a resolution of 601 × 601 pixels, a focal length of 10mm, and an interocular distance of 6cm. We repeated the calculation for 100 different vantage points, corresponding to different positions and orientations of the cyclopean visual axis, and for a set of fixation points. The fixation points varied in the range of 0◦ ÷360° for the azimuth angle, and in the range of 0° ÷32° for the polar angle. More precisely, the fixation points were obtained by backprojecting a 11×11 grid of equally spaced points of the cyclopean retina on the closest visible surface of the scene. Under the same experimental conditions, the disparity patterns were calculated for three different eye movement paradigms (Tilt-Pan, Listing and its binocular extension L2). The mean vector disparity patterns, together with their standard ellipses (measuring the joint

dispersion of the bivariate distribution) are shown in Fig. 12E. It is worth noting that, although the mean disparity patterns calculated for each fixation are characterized – as expected – by significant differences, these differences are attenuated when averaging over all the fixations we considered (not shown).

Empirical corresponding points and the reference surface – The horopter, as the locus of point in the 3D space whose projections fall on geometrically corresponding points in the two retinas, usually consists of two parts: the horizontal and the vertical horopter. The former lies in the horizontal plane of regard, and it is described by the Vieth-Müller circle, *i.e.*, a circle through the nodal point of the the two eyes and the fixation point. The latter is a line through the Vieth-Müller circle, in the median plane of the head, and at right angle to the horizon plane.
Together with the geometrically corresponding points it is possible to define a second type of correspondences on the basis of psychophysical or physiological criteria, such as singleness of vision or nonius alignment: the empirical corresponding points. Several researchers tried to measure the pattern of empirical corresponding points, and they all agree that these do not coincide with the geometric ones (Schreiber *et al.*, 2008). More precisely, they claimed that empirical correspondences are not geometrically congruent along the horizontal meridian, and that points in the temporal hemiretinas are relatively compressed with respect to the corresponding points in the nasal hemiretinas. Relative to the geometric points, the empirical points have negative (i.e. uncrossed) disparities to the left and to the right of fixation. This compression causes the horizontal horopter to deviate from the Vieth-Müller circle. This difference is known as the Hering-Hillebrand deviation (Howard & Rogers, 2002). The horizontal horopter tends to be less concave than the Vieth-Müller circle at near distances and more convex at far distances. Moreover, empirical corresponding points are anisotropic. This anisotropy is illustrated by the fact that corresponding vertical meridians are sheared horizontally. This means that empirical correspondences are characterized by negative (i.e. uncrossed) disparities in the upper part of the retina, above the fovea, and by positive (i.e. crossed) disparities in the lower part, whose magnitude increases with the eccentricity. This anisotropy is called Helmholtz shear deviation, and it causes the vertical horopter to incline top away in the median plane (Schreiber *et al.*, 2008).
From this perspective, we can consider the mean disparity patterns, obtained friom the statistical analysis, as a pattern of empirical corresponding points. From this it is possible to derive a "minimum-disparity" horopter as the 3D surface whose projections have the minimum disparity angle with the pairs of empirical corresponding points. The resulting optimal surface is a tilted top away surface at a distance of 104cm, less concave than the Vieth-Müller circle, and its concavity decreases with the fixation distance, in agreement with the experimental observations, see Fig. 13.

Task 1.3: ***Control of voluntary eye movements in 3D***
This Task is devoted to the study of ocular mechanics and oculomotor control, for both single eye and conjugate movements. The target is to investigate how eye plant mechanics affects the strategies implemented by the brain to drive typical biological ocular motions (including saccades and smooth pursuit). A second goal is the study of the geometric and kinematic effects of ocular motions on image flows, for supporting the estimation of 3D information from ocular motions. Finally, from the engineering point of view the major expected achievement is to provide the guidelines for the development of a bio-inspired stereoscopic robot system capable of emulating the ocular motions.
In order to meet these objectives it has been necessary to develop a detailed model of the ocular mechanics and of the control strategy (at muscolar level) that realizes the human eye movements in the 3D space. A detailed mechanical model of the oculomotor system (i.e., eyeball and extraocular muscles) has been developed, and its geometrical and dynamical properties against Listing's Law have been described.
The major outcome of this activity has been a complete 3D dynamic model describing the Listing's Law dynamics featuring a (simplified) model of the visco-elastic featured of the orbital tissue surrounding the eyeball. However, even though the proposed model is compatible with Listing's Law, so that for any action force generated by the rectii extra ocular muscles (EOMs) the eye orientation has zero torsion (Listing's

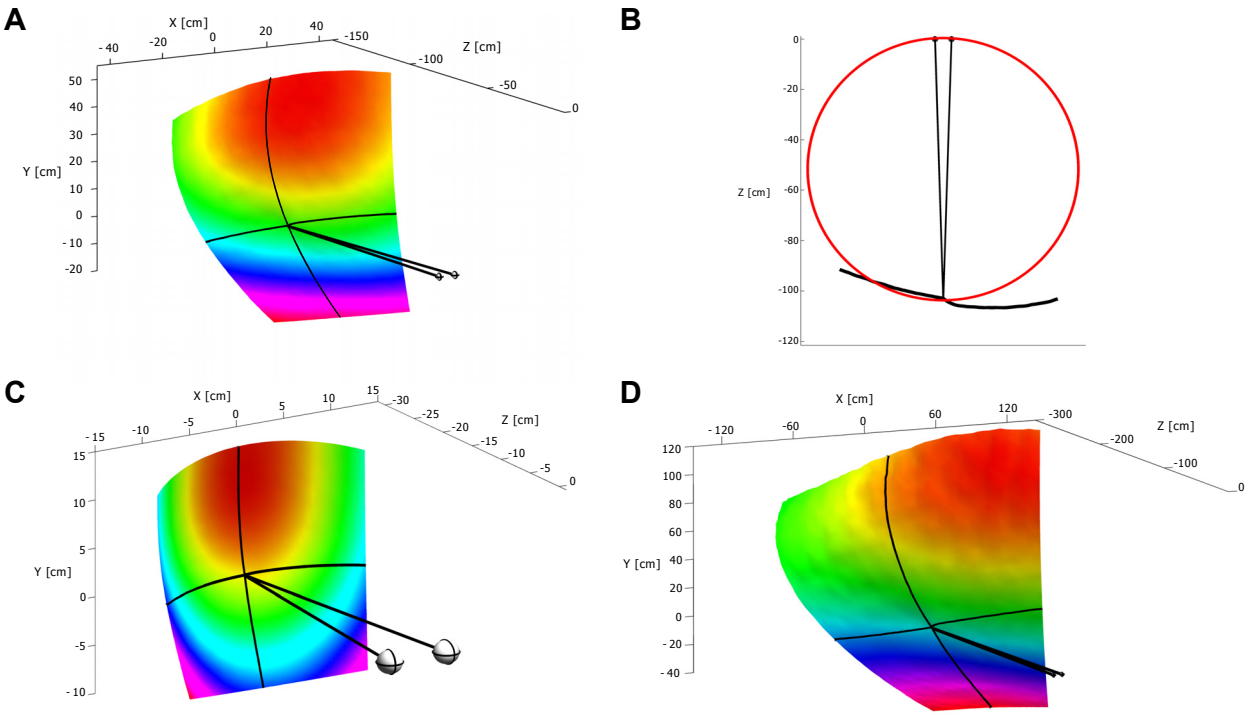


Figure 13: *The empirical horopter for different fixation distances.* ***(A)*** *The optimal surface. For a fixation distance near 104cm we found the surface that minimizes the disparity between the stimulated points on the retinas and the empirical corresponding points. The heavy black lines represent the horizontal and the vertical components of the horopter.* ***(B)*** *The horizontal component of the optimal surface depicted in (A) (thick black line) compared with the Vieth-M¨uller circle (red line). The fixation distance acts on the empirical horopter changing its concavity and its tilting. At 30cm* ***(C)*** *the horopter is more concave than in (A), whereas it becomes flatter at 200cm* ***(D)****.*

Law), still remains the problem of computing the appropriate action forces required to rotate the eye to a given target direction.

As a matter of fact the high dynamic response characteristics of the saccadic motions pose a significant challenge for implementing a closed loop control strategy. Furthermore, in a bio-inspired framework, long latencies featuring the vision system and the lack of evidence of mechano receptors dedicated to measure the posture of the eye, make the design of high performance closed loop strategies an issue. As a matter of fact, open loop time optimal control strategies have been proposed in the literature to explain how the actual control EOMs' control actions could be computed to generate saccadic motions (Clark & Stark, 1975; Enderle, 1984). Experiments (Wang *et al.*, 2007) have shown that in monkeys there exists a neural representation of the eye orientation that could be based on muscular stretch receptors. Although these signals are not suitable for high bandwidth feedback control, they can be used for fine adjustment of eye position or recalibration of the proprioceptive system (Wang *et al.*, 2007).

To tackle this control problem we have addressed two basic issues that play a relevant role in the problem of controlling the eye orientation. The first one is how to compute the steady state action forces generated by the EOMs to keep the fixation in a given target direction. The second one is instead related to the problem to compute, at steady state, for a given set of EOMs' actions the eye orientation. These two problems correspond to static inverse and direct problems and play a complementary role. We will assume that some sort of force feedback from the EOMs is available.

The Static Direct Problem (SDP) correspond to a computational procedure for estimating the eye orientation given the actual action forces generated by the EOMs and could be used a low bandwidth signal, and provides a map associating the EOMs' actions to the eye orientation.

The Static Inverse Problem (SIP) is instead required to compute the action forces of the EOMs to be generated at steady state in order to keep the orientation of the eye to some target value. This is an ill-posed problem as there are four the three rotational degrees of freedom of the eye are controlled by four EOMs, so

there are infinitely values of the motor commands that correspond to a unique eye position. For example, in static condition, the tensions of the muscles could be proportionally increased, so that the total torque does not change, thus leaving the eye position unchanged.

The SDP problem

The Static Direct Problem is the procedure required to compute the steady state eye orientation given a constant set of action forces generated by the EOMs. To solve this problem we have proposed an iterative procedure (Cannata & Trabucco, 2011 [C17]), validated using the Simulink based simulator developed in Task 1.4.

The SIP problem

Here, we investigate the problem of associating the action forces of the EOMs required to mantain the eyeball in a given orientation. Figures 14-15-16-17 highlight the controlled rotations of 20^o about the vertical axis ($\alpha = 90^o$) and about a vector oriented at $\alpha = 60^o$ with respect to the horizontal line (and belonging to the Listing's Plane). As it is apparent, the resulting motion is accurate and respects all the motion constraints.

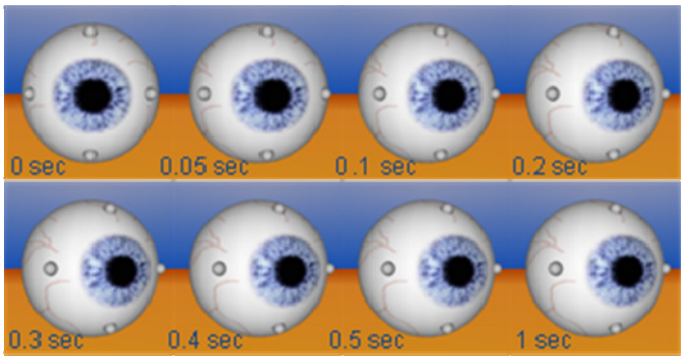


Figure 14: *Simulated eye rotation with constant forces computed using the SIP algorithm (case $\alpha = 90^o$).*

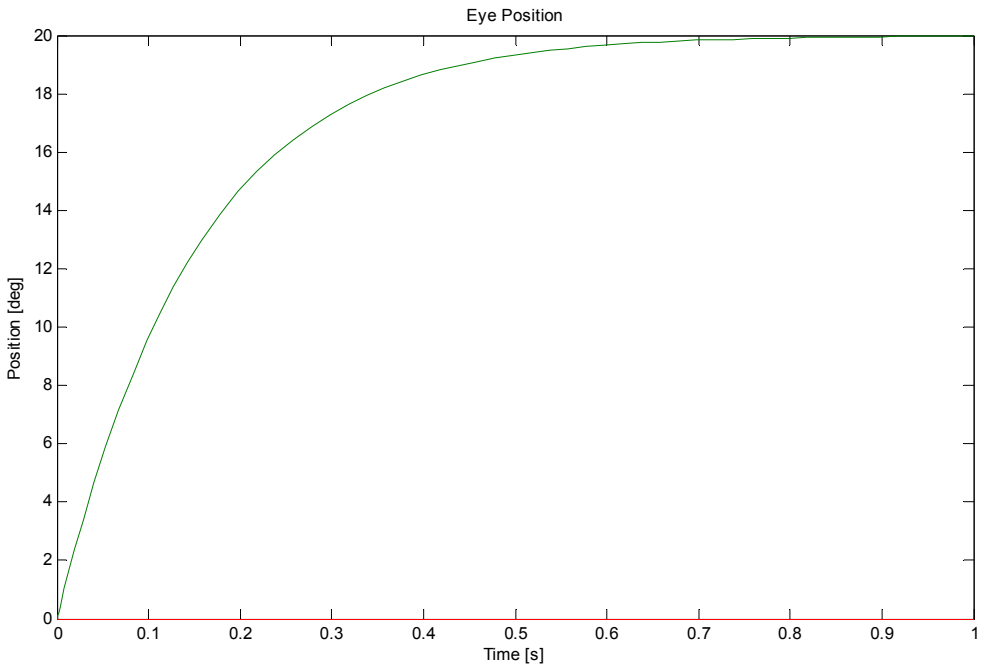


Figure 15: *The eye position trajectory with constant forces computed using the SIP algorithm (case $\alpha = 90^o$).*

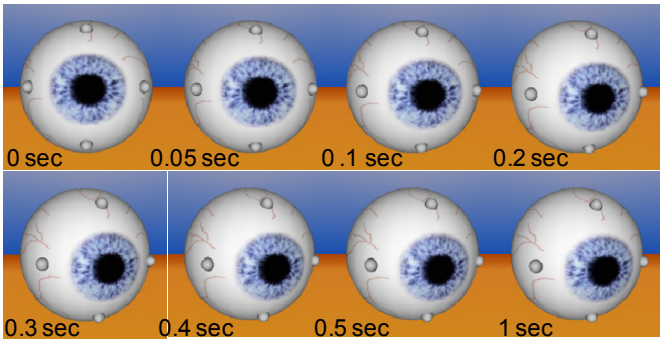


Figure 16: *Simulated eye rotation with constant forces computed using the SIP algorithm (case α = 60°).*

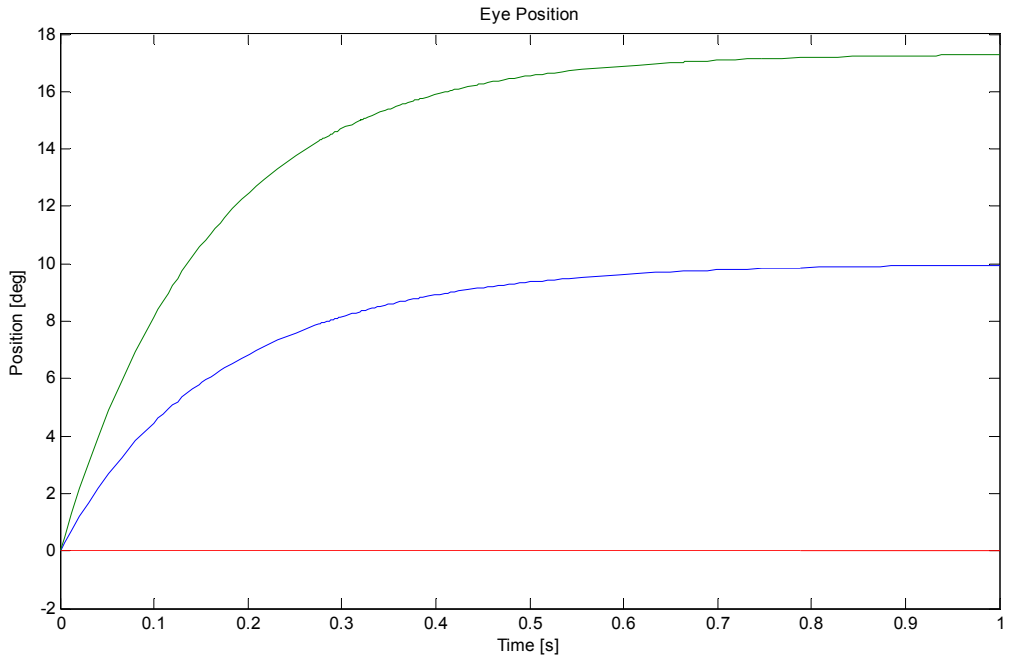


Figure 17: *The eye position trajectory with constant forces computed using the SIP algorithm (case α = 60°).*

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

Many eye-head robots have been developed in the past few years, and several of these are common pan-tilt systems, where a camera rotates about pan-tilt axes. The main goals of this work have been to provide the guidelines for the implementation of a tendon driven robot and to emulate the different types of human eye movements for binocular vision experiments.

The geometry of the robot prototype is based on the model described in Task 1.3 and a solution to emulate the mechanical properties (viscosity and elasticity) of the eye orbital soft tissues have been considered.

Relevant features are also the dimensions of the robot, which are very close to the human eye. The final dimensions have been the result of a careful condiseration of the various trade-offs during the selection of the components available on-the-shelf (e.g. the eyeball, motors, on board camera etc).

The relevant subsystems are: the eyeball and the vision system; the support structure; the actuation system.

Robot Eye Design

This prototype of a robot eye must emulate the mechanical structure and the movements of a human eye with a comparable working range. This robot eye has approximately the shape of a truncated cone where the larger diameter and the smaller one are 34 and 45 mm, respectively, with an overall length of 120 mm. The robot is designed on the assumptions that the eyeball is a sphere with three degrees of freedom about its center and the actuation system that drives the eyeball is a combination of linear motors, springs and tendons.

Each linear motor has two springs in parallel, the motor and the springs are connected to the eyeball through the tendons. Figures 18 and 19 show the three dimensional CAD model and a detailed view of the system.

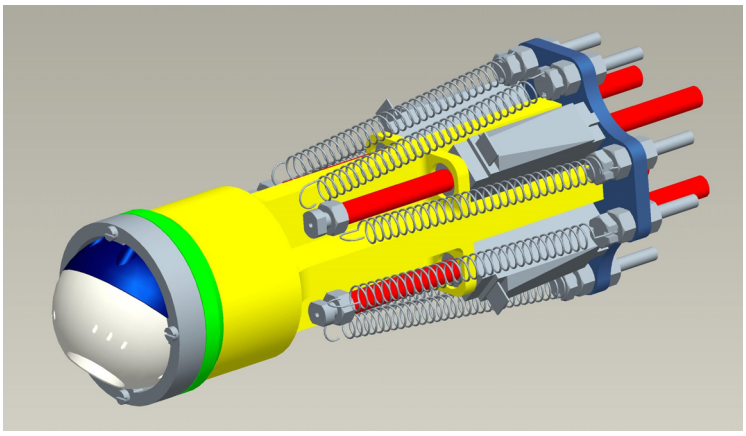


Figure 18: *3D sketch of the robot eye prototype.*

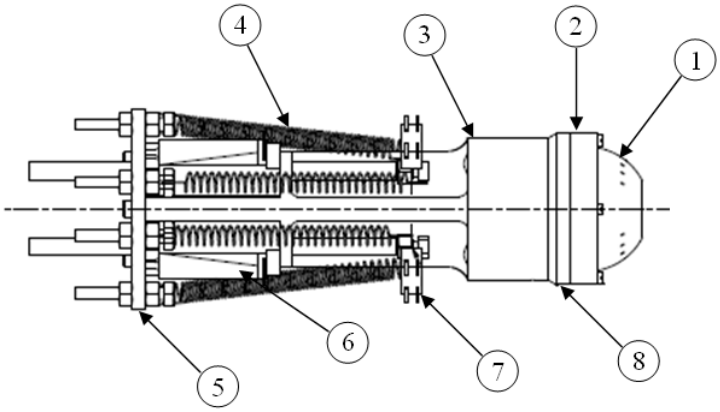


Figure 19: *Lateral view of the eye robot prototype and its modules:* ***1)*** *Eyeball,* ***2)*** *Front flange,* ***3)*** *Frame,* ***4)*** *Spring,* ***5)*** *Rear flange,* ***6)*** *Linear motor,* ***7)*** *Position sensor,* ***8)*** *Eyeball support.*

The Eyeball

The eyeball is a precision machined (in house) PTFE sphere with a diameter of 28 mm. The sphere has been machined to host the vision system (a commercial CMOS microcamera with miniature optics) and to route the video signal cables to the external electronics.

Supporting structure

The structure designed to support the eyeball, the motors and the springs is composed of four distinct components (shown in Fig. 20 and described below):

- eyeball support: a low friction support designed to hold the eyeball and to implement the pointwise pulleys,
- frame: a structure cross, attached to the eyeball support, to hold the motors and the springs,
- front flange: an anterior flange to lock the eyeball on the eyeball support,
- rear flange: a posterior flange to lock the motors and the springs on the frame.

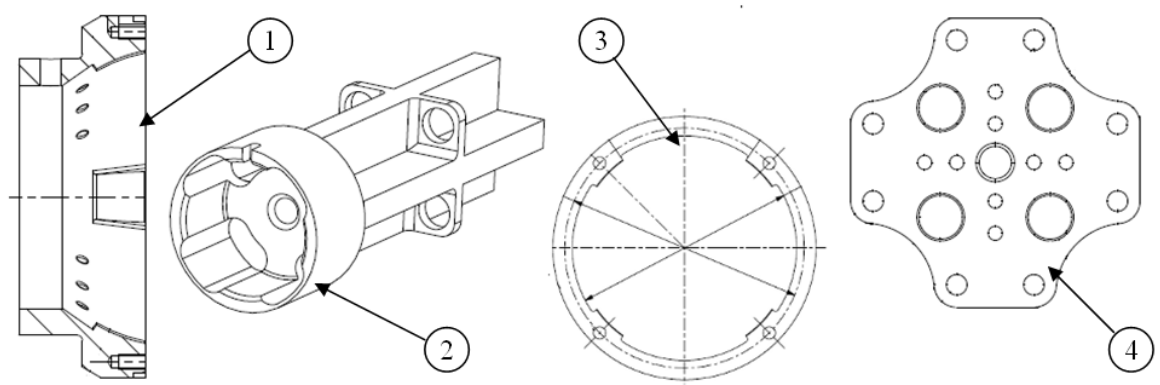


Figure 20: *CAD model of the supporting structure:* ***1)*** *Eyeball support,* ***2)*** *Frame,* ***3)*** *Front flange,* ***4)*** *Rear flange.*

The eyeball support module holds the eyeball and is the most critical part of the system. It is made of TEFLON and it has the major function of implementing the pointwise pulleys (responsible to guarantee mechanical implementation of Listing's Law). The pulleys route the actuation tendons and they ensure the correct mechanical implementation of Listing's Law.
The position of the pointwise pulleys is symmetrical with respect to the position of the insertion points on the eyeball. On the eyeball support there are four groups of three pulleys: the central for the tendon attached to the linear motor and the two lateral for the tendons attached to the springs.
Figure 21 shows a sketch of the posterior and two different lateral views of the eyeball support.

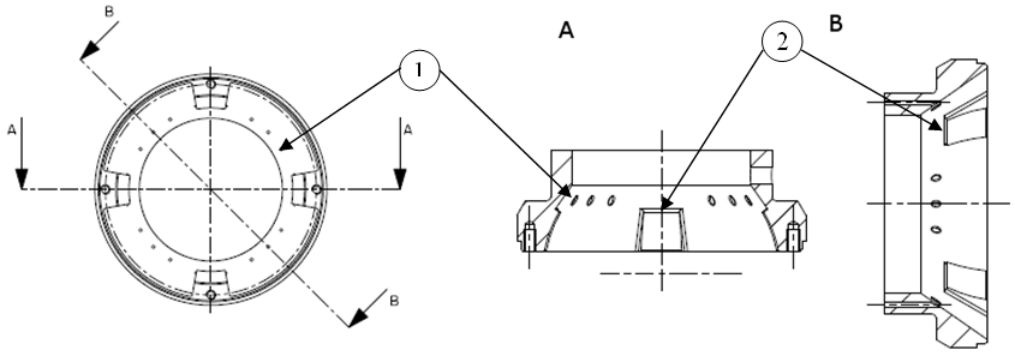


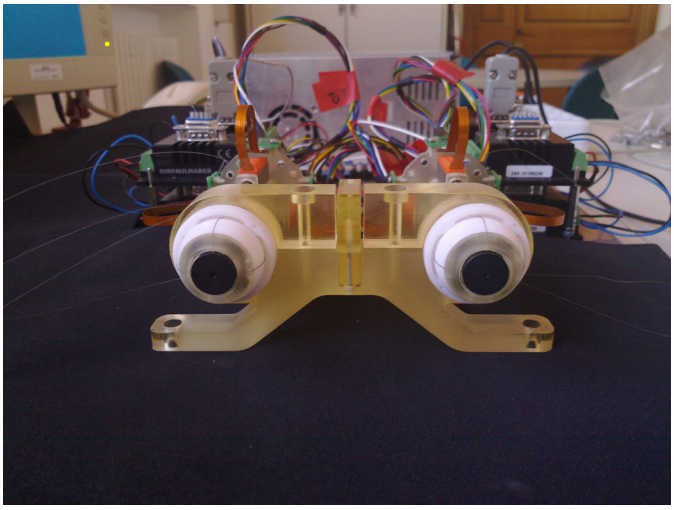
Figure 21: *CAD model of the eyeball support: posterior and two lateral views* ***(A,B)****: 1 Pointwise pulley holes, 2 Screw holes.*

Actuation System
The actuation system is composed of four tendons, four force generators (DC brushless linear motors) and eight springs. The tendons are thin stiff wires, connected to the rods of the motors and to the springs. The actuators pull the tendons and drive the movements of the eyeball. There are two springs in parallel to each motor that pull the tendons in the opposite direction with respect to the motor one. The main function of the springs is to emulate the elasticity of the orbit of the human eye and to restore the zero position (primary position) of the eyeball when the system is not actuated. Different views of the current version of the robot prototype are shown in Fig. 22.

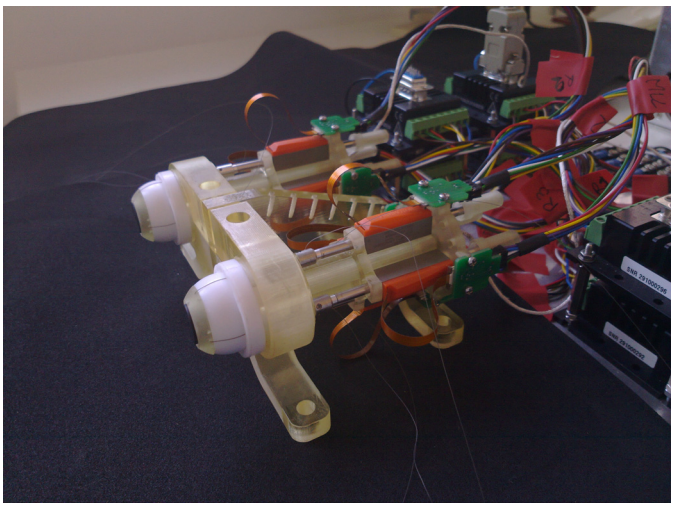
Deviations from the project workprogramme

The development of the simulator, planned after the First Review Meeting (April 2009) has delayed by almost one year the development of the robot prototype. In order to keep the development of the robot eye prototype within the project life span *non-conventional* manufacturing solutions have been adopted (plastic rapid prototyping). The risks of this approach are high (due to the lower accuracy and limited resistance of the prototypes), however, careful redesign of the critical mechanical parts has led to a satisfactory working prototype.

Front view



Side view



Top view

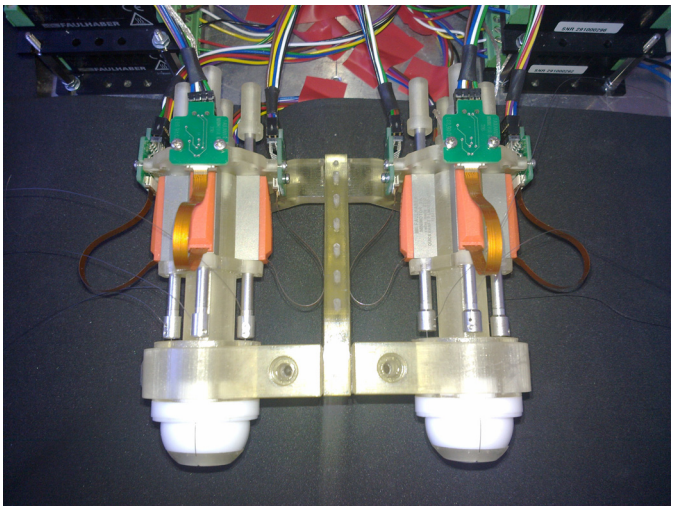


Figure 22: *Pictures of the first prototype of the tendon-drive binocular robot head.*

WP2: Active stereopsis

Leader: Marc Van Hulle (K.U.Leuven)
Contributors and planned/actual effort (PMs) per participant:
UG (7/4.55) and K.U.Leuven (17/18)
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 real robotic heads, due to the limited accuracy of their motor system. Therefore, vision is used to improve upon this. In the previous period we had developed, and provided software, for autocalibration methods that can operate in the retinal as well as in the cortical domain. Certain aspects of these methods are difficult to align with the experimental evidences reported in various neurophysiological studies. Therefore, we have now applied these same principles to develop a biologically plausible architecture.

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, the research activities on (1) the functional specialization of cortical-like disparity detectors to implement a 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.
In the third period, we have extended our simulation efforts and spend more attention on getting statistically quantifiable results (mean values and variances) on the performance of the proposed control. Different vergence maintenance experimental tests have been carried out distinguishing arbitrary fixations on smooth surfaces and arbitrary fixations in presence of depth discontinuities. On this basis, we improved the networks' performance in order make them properly work in real-world scenarios on robotic platforms. In addition, limited to the dual-mode vergence control model, we demonstrated that the disparity-vergence responses can be steered to cope with the distortions of the epipolar geometry that occur in space-variant image sensing schemes, with the advantage of extending the working range of the control without requiring additional computational resources.

Step 1: Linear/non-linear convolutional networks for learning vergence control

For the vergence control paradigm modeling, we have used the framework shown in Fig. 23. This setup consists of the vergence simulator module, the disparity detector population module, the population response post-processing module and the vergence control network (VC-net) module.
The main goal of the vergence simulator is to generate a stereo image (left and right eye views) based on the actual state of the robotic head: the vergence angle and the gaze direction, and information about the 3D environment.

Vergence database – For training the VC-net, we have prepared a vergence database. The database consists of two tables: a table of synthetic scenes and a table of vergence samples (see Fig. 24). For efficient memory usage, the scenes were allowed to be reused in several vergence samples. There are two types of synthetic scenes in the vergence database, which correspond to the simplified- and general case scenarios, respectively. The simplified case scenes contain only one type of object-stimulus, a fronto-parallel rectangular patch perpendicular to the gaze direction in the primary position. The stimulus in this case is large enough to completely cover the field of view of both cameras.

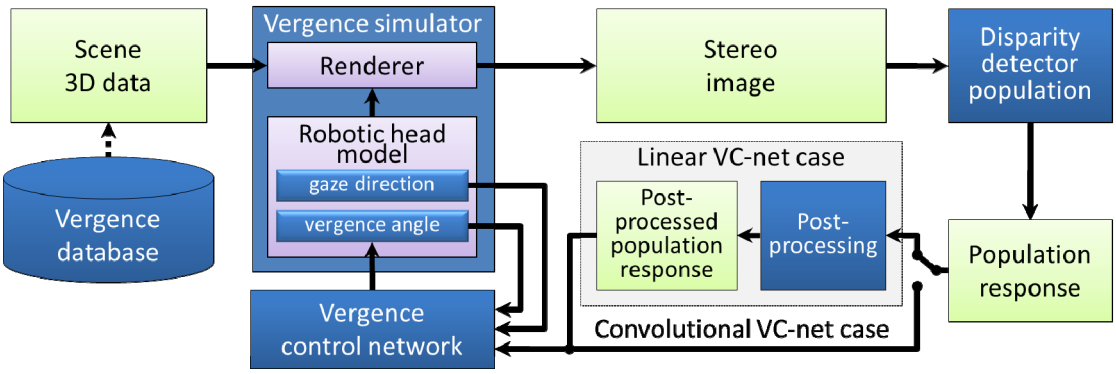


Figure 23: *The block diagram of the framework used in vergence control model training and testing. The stereo image generated by the simulator is processed by the disparity detector population, to produce the population response. Depending on which vergence control network is used, the population response is then directed to either the population response post-processing block, which is producing the post-processed population response (the linear VC-net case), or directly to the vergence control network module (the convolutional VC-net case). The (raw/post-processed) population response, together with the actual values of the gaze direction and the vergence angle, are fed into the vergence control network module, the main module of the model. The goal of the VC-net is to produce a new vergence angle, to get the fixation point onto the surface of the object of interest, without changing the gaze direction.*

The general case scenes consist of several simple textured objects, randomly placed into a room-like virtual environment with several light sources. The object sizes are chosen randomly allowing for depth discontinuities. Vergence samples consist of the gaze direction, the actual vergence angle, the stereo pair (left and right eyes' images), the population response for the stereo pair and the desired vergence angle. The actual vergence angle is a perturbed (with Gaussian noise) version of the desired one. The actual vergence angle is expected to become as close to the desired vergence angle as possible, when running the control model. Each vergence sample in the database can be considered as a training pair. The input part is constructed from the post-processed (or raw) population response, the gaze direction and the actual vergence angle; the output consists of only one scalar parameter (the desired vergence angle). The vergence database used for the VC-net training consists of 1000 synthetic scenes and 5000 samples. The balance between general and simplified scenes (as well as for the samples) has been set to 50/50%. Real-world images were used as textures for the objects. To reduce the influence of a possible overfitting to particular textures on the results of the evaluation, we have used non-overlapping sets of textures for the training- and test experiments. An early stopping technique (with 10% of the training data for validation) was used to prevent overfitting during training. To achieve a fair comparison, both VC-nets were trained using the same training data.

Vergence simulator – The vergence simulator module consists of the renderer and the ideal robotic head model (RHM) with fixed neck. In this model, the robotic head is assumed to be fixed and the eyes to rotate around their nodal points. We selected this model because it is easy to implement, and eventually to replace by a real tilt-pan stereo setup.

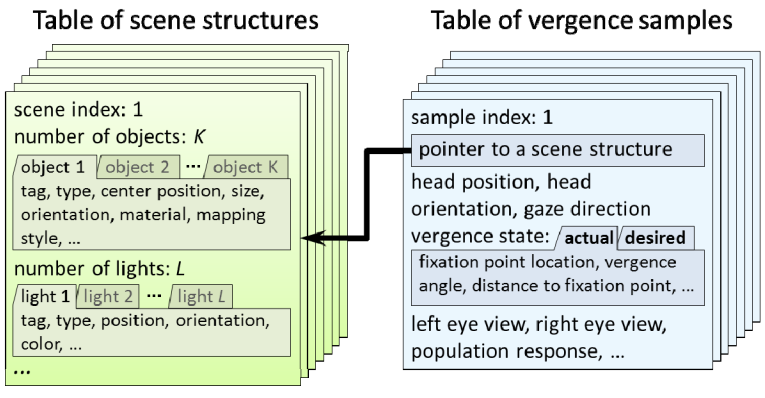


Figure 24: *Schematic structure of the vergence database.*

The renderer, in turn, produces the stereo image, observed by the left and right eyes, using the position/orientation of the eyes, and the geometric description of the scene, provided by the scene 3D data block. To make sure that the disparities are not too large and can be properly handled by the disparity detectors, we decided to render the retinal projections with low resolution i.e., we obtain images of 41×41 pixels for a field of view of 20°. Note that the resolution could be higher, but consequently to allow the population to cope with the same range of disparities, the receptive fields of the disparity detectors should be larger, which would significantly increase the computational cost and, thus, slow down the simulations.

Disparity detectors population module – Disparity information is extracted from a stereo image pair through a distributed cortical architecture discussed in great detail in previous reports. We considered only a single-scale disparity detector population, but the population can be readily extended to the multiscale mode, without conceptually changing our framework, but which will be computationally much more expensive.

Post-processing module – The post-processing of the population response is used only for the linear VC-net, and comprises a two-dimensional convolution over the first two (spatial) dimensions of the population response, using a two-dimensional Gaussian kernel G_σ:

$$P_{ij} = G_\sigma * r_c^{ij},$$

where r_c^{ij} is the population response map for the i-th orientation and the j-th phase shift. The kernel G_σ has the same size $n_r \times n_c$ as the size of a population response map r_c^{ij}, so the result of the convolution is a scalar value P_{ij}. On the one hand, this step drastically reduces the amount of data to further process. Indeed, after pooling, the network has to process only a 2D ($N_o \times N_p$) pooled population response instead of a 4D ($n_r \times n_c \times N_o \times N_p$) array, where N_p is the number of phase shifts, and N_o the number of orientations. But, on the other hand, the pooling has a major drawback as it discards the spatial information about the disparity encoded in the population response. The results of simulations revealed that, in the general case scenario, this discarding could lead to degraded vergence accuracy. The convolutional network works directly on the population response, and the post-processing is done in the first two layers of the convolutional network.

Vergence control module – This module is the main module of the model. The purpose of it is to convert the post-processed population response together with the actual vergence, and the gaze direction, into a new vergence angle. Virtually, this module can be represented by any kind of paradigm, but in this workpackage we discuss only a linear network and a convolutional network.

- *Linear networks*: the simplest possible solution consisting of only a single linear unit that collects the weighted sum of the population responses. The simulations revealed that even this simple network is able to produce accurate angular vergence control in some restricted situations (*e.g.*, in the simplified case). The input vector for the linear VC-net was constructed as a concatenation of the pooled population response (56 values), the gaze direction (2 values) and the actual vergence (1 value), so its dimensionality is 59. The output is a prediction of the vergence angle, which is a scalar value (see Fig. 25). Due to the linearity of the network, there was no reason to introduce any hidden layers, so the linear VC-net consisted of only one linear unit. This simplest possible vergence control network has only 60 parameters (including bias), which can be learned either directly (using linear regression or its robust modification), or iteratively (using gradient descent), from the training database.

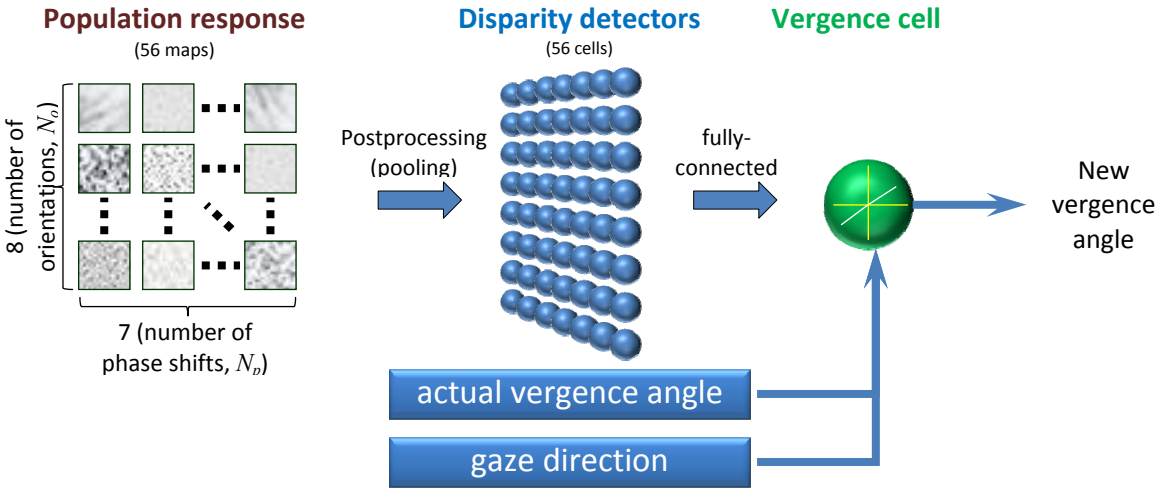


Figure 25: *Linear vergence control network and its inputs.*

- *Convolutional networks* (CNs): A typical convolutional network is a feed-forward network of layers of three types: convolutional (C-layer), subsampling (S-layer) and fully-connected (F-layer). The C-layers and S-layers usually come in pairs and are interleaved, and F-layers come at the end. The output of a C-layer is organized as a set of feature maps. Each feature map contains the output of a set of neurons with local receptive fields. All neurons in the feature map share the same weights, so the feature map is responsible for a particular local visual feature, encoded in the weights of these neurons. The computation of a feature map starts with a 2D convolution of the input with a fixed kernel defined by the neuron's weights. A feature map can have inputs from several feature maps of the previous layer. In order to condense the extracted features, and to make them more invariant with respect to spatial deformations, the C-layer is typically followed by an S-layer which performs a local averaging and subsampling. Each neuron in F-layer just adds a bias to the weighted sum of all inputs and then propagates the result through a nonlinear transfer function (RBF or sigmoid).
  The network is trained in a supervised manner using backpropagation. For the efficient training of large CNs, LeCun and colleagues proposed a modification of the Levenberg-Marquardt algorithm (LeCun, 1998).
  The architecture of the convolutional network, used for our experiments is shown in Fig. 26. The main challenge in this approach was the amount of data: the population response consists of 56 (8×7) maps of resolution 41×41 (rendered image resolution), so the input of the network has 94136 (41×41×8×7) components. In order to be able to train the network with such high dimensional input data, we had to reduce the number of training parameters. The first (convolutional) layer is a fixed set of (nontrainable) Gaussian kernels of size 19×19 with standard deviation 6. The second (subsampling) layer has also 56 feature maps size of which was set to 3×3.

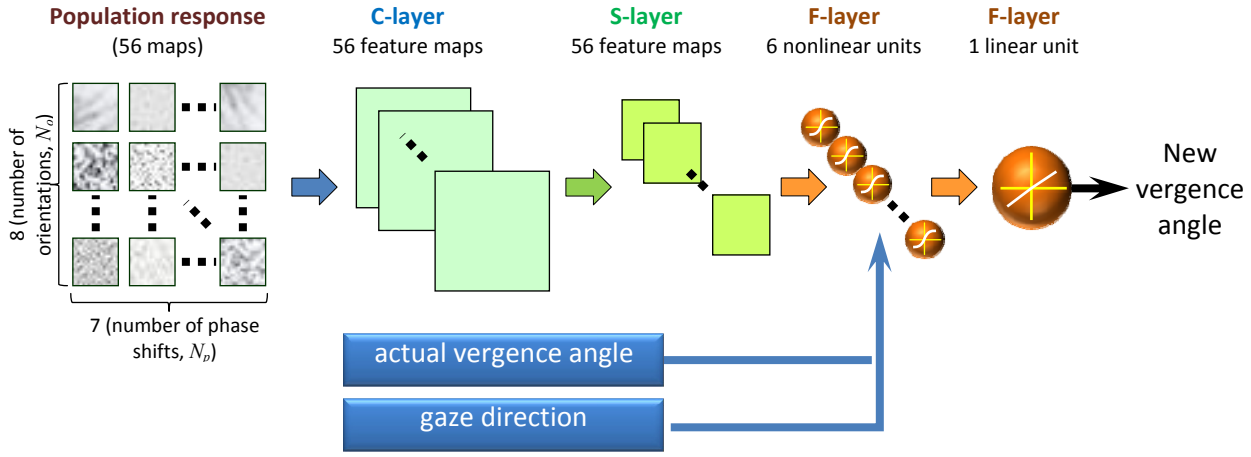
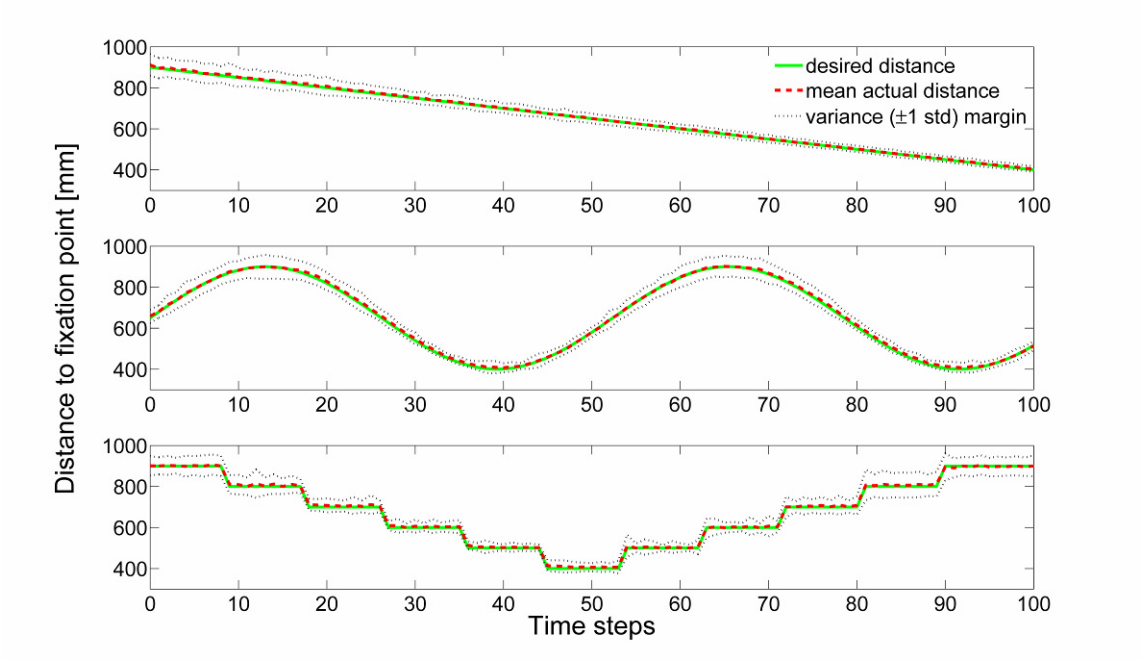


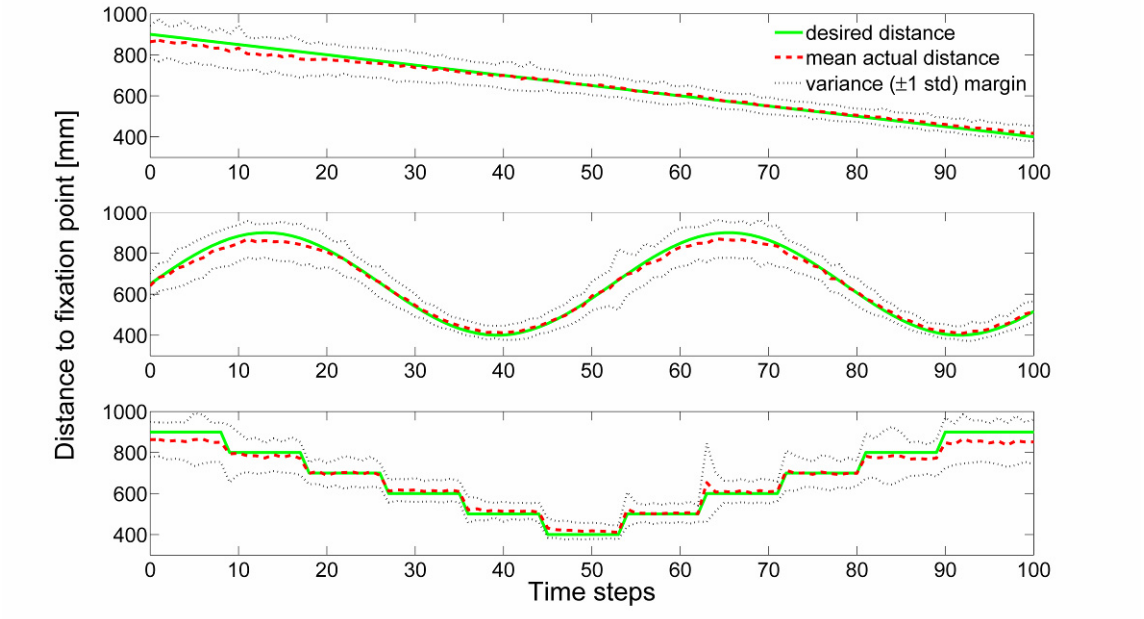
Figure 26: *Convolutional vergence control network and its inputs.*

Results – To evaluate both VC-nets a series of 100 vergence maintenance experiments have been carried out for both the simplified and a the general scenario. Each experiment consisted of 100 steps during which the randomly generated stimulus was moving along the gaze direction, changing its distance (from 400 mm to 900 mm) to the head in a particular manner. We have considered three patterns of the stimulus motion-in-depth: ramp, sinusoid and staircase. Pre-trained VC-nets were allowed to control the actual vergence angle to keep the fixation point as best as possible on the surface of the stimulus. During each experiment, the actual and the desired values of the vergence angle, and the distance to the stimulus were stored for each time step, for further analysis.
The results of the evaluation of above mentioned experiments of both VC-networks in both considered scenarios are presented in Fig. 27, Fig. 28 and Table 1. Each panel of Figs. 27 and 28 contains: 1) the desired (ground truth) distance to the stimulus curve depicted by the solid green curve, 2) the mean (averaged across all experiments) actual distance to the stimulus curve depicted by the dashed red curve, and 3) the variance (standard deviation across all experiments) of the actual distance margins depicted by the dotted black curve.
The performance of the VC-net can also be assessed using the ratio of the distance-based error variance to the corresponding desired distance. The smaller this ratio is, the lower the relative (distance) error is produced by the network. Table 1 contains the minimum, mean, median and maximum values of this ratio (in percent) for each experiment type and each stimulus.

From Fig. 27 and Table 1, it can be clearly seen that both networks perform relatively well in the simplified scenario: the mean actual distance curve almost coincides with the desired one, and the variance in both cases is relatively small. For the general case scenario, the situation is different. The linear VC-net (Fig. 27b) shows a much larger variance and a general tendency to over(under)shoot towards the "average" depth of the scene (at approximately 600 mm). The convolutional VC-net (Fig. 28b) also shows a relatively larger variance, but the mean actual distance is closer to the ground truth than in the linear VC-net case. The effect of the anisotropy of the distance uncertainty, mentioned previously, is noticeable in Fig. 27 and Fig. 28: the further the stimulus is, the larger are the mistakes made by the VC-net.

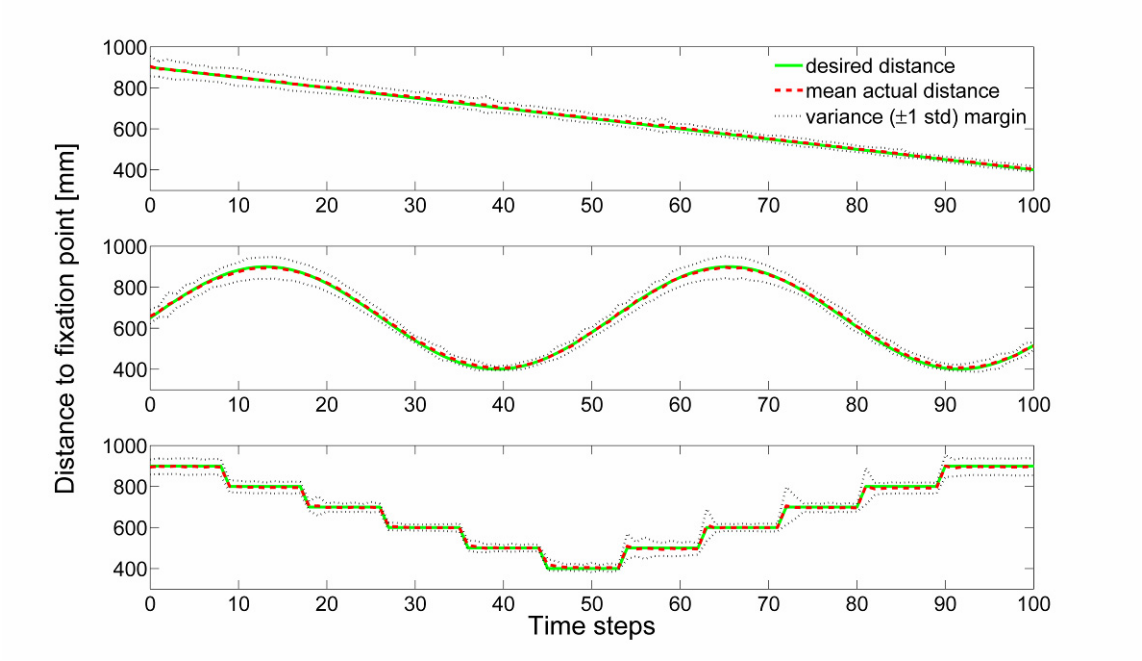


(a) Linear VC-net, simplified scenario.

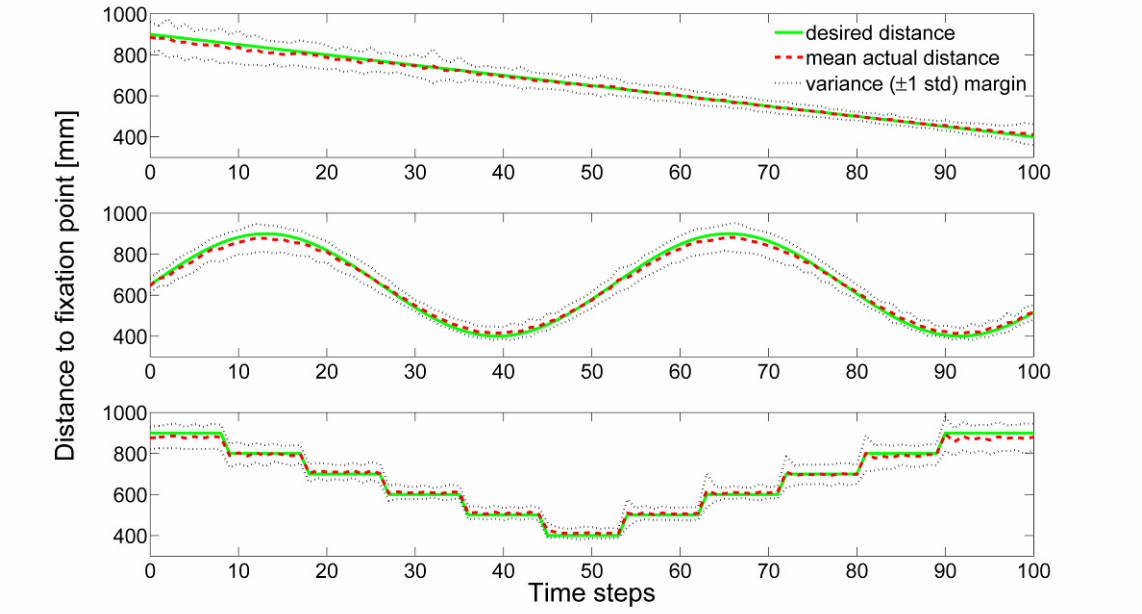


(b) Linear VC-net, general case scenario.

Figure 27: *Results of the depth-based performance plots for linear VC-net in both scenarios.*



(a) Convolutional VC-net, simplified scenario.



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

Figure 28: *Results of the depth-based performance plots for convolutional VC-net in both scenarios.*

Table 1: *Variance of distance-based error relatively to desired distance.*

VC-net	Experiment scenario	Stimulus type	Error variance ratio (%)			
			min	mean	median	max
Linear	Simplified case	Ramp	2.6828	3.8921	3.6172	6.2715
		Sinusoid	2.8904	5.2275	5.2840	7.5772
		Staircase	2.6566	5.4345	5.2589	10.6765
	General case	Ramp	6.4996	8.4466	8.1057	12.9288
		Sinusoid	6.2622	10.0322	9.9448	22.1558
		Staircase	7.1387	10.6848	9.9420	31.9260
Convolutional	Simplified case	Ramp	2.4841	3.7045	3.6237	5.8980
		Sinusoid	2.2913	4.8870	4.9034	8.6034
		Staircase	2.1722	4.3578	3.6502	13.2121
	General case	Ramp	4.0339	6.4378	5.7930	12.7739
		Sinusoid	4.8622	6.9828	6.8680	10.6242
		Staircase	3.9617	6.5304	6.2880	13.7682

Step 2: Impact of space-variant log-polar image sensing on the dual-mode vergence control

In designing an active control for a robot stereo head, it is necessary to take into account the large amount of visual information that mus be processed in real time. According to the dual-mode vergence model developed by partner UG (Gibaldi et al., 2010a), the left and right images are processed by a population of disparity detectors, inspired by complex cells of area V1. The population provides a distributed representation of the retinal disparity, and through convolutions with weighting kernels it is decoded to obtain a family of vergence cells that can drive a direct vergence motor response, in accordance with the neurophysiological evidences in area MST. Since the task is to drive vergence eye movements so as to improve the fixation and thus the estimation of disparity, the necessary disparity information is gathered only from the central (perifoveal) portion of the visual field (see Fig 29A).

Since a real-time behavior is necessary for a stable and effective control, specific design strategies must be considered to reduce the computational burden associated to image processing. Taking inspiration from the mammalian visual system, it is possible to implement a space-variant mapping of the stereo images in order to simplify the computational problem for active vision (Traver & Bernardino, 2010). Indeed, the retinal receptive fields show a high spatial resolution and a small extent in the central part of the retina, the fovea, and a larger extent that increases with the eccentricity (Schwartz, 1977). Such a transformation is well described by a log-polar mapping form the retinal (Cartesian) domain to the cortical one that preserves a high resolution in the area of interest, and a compression of the information in the periphery. We extended the dual-mode cortical model for the control of vergence to work with space-variant resolution images, arranging the resources according to a log-polar geometry.

To "optimally" design the log-polar mapping for visual processing tasks, it is important to study the relationships between the usual processing in the retinal domain and the direct extraction of the features in the cortical domain, by characterizing the filters with respect to the different parameters of the log-polar mapping. To solve the problem of the singularity of the fovea, we used the "blind-spot" model (Traver and Pla, 2008). In order to allow the binocular energy model to work with log-polar images, we considered that the space-variant geometry, by compressing the information in the periphery, produces a distortion of the image, and consequently of the disparity information in the binocular image. Following (Solari *et al.*, 2011 [J6]), we designed the transformation in order to have a minimal distortion of the information, at least locally.

Besides *local* distortions, the log-polar mapping introduces, at a *global* scale, deformations of the epipolar geometry that transforms horizontal disparity in the retinal domain (see Fig. 29C, left dashed blue lines), in space-variant oriented disparities the cortical domain (see Fig. 29C, right dashed blue lines).

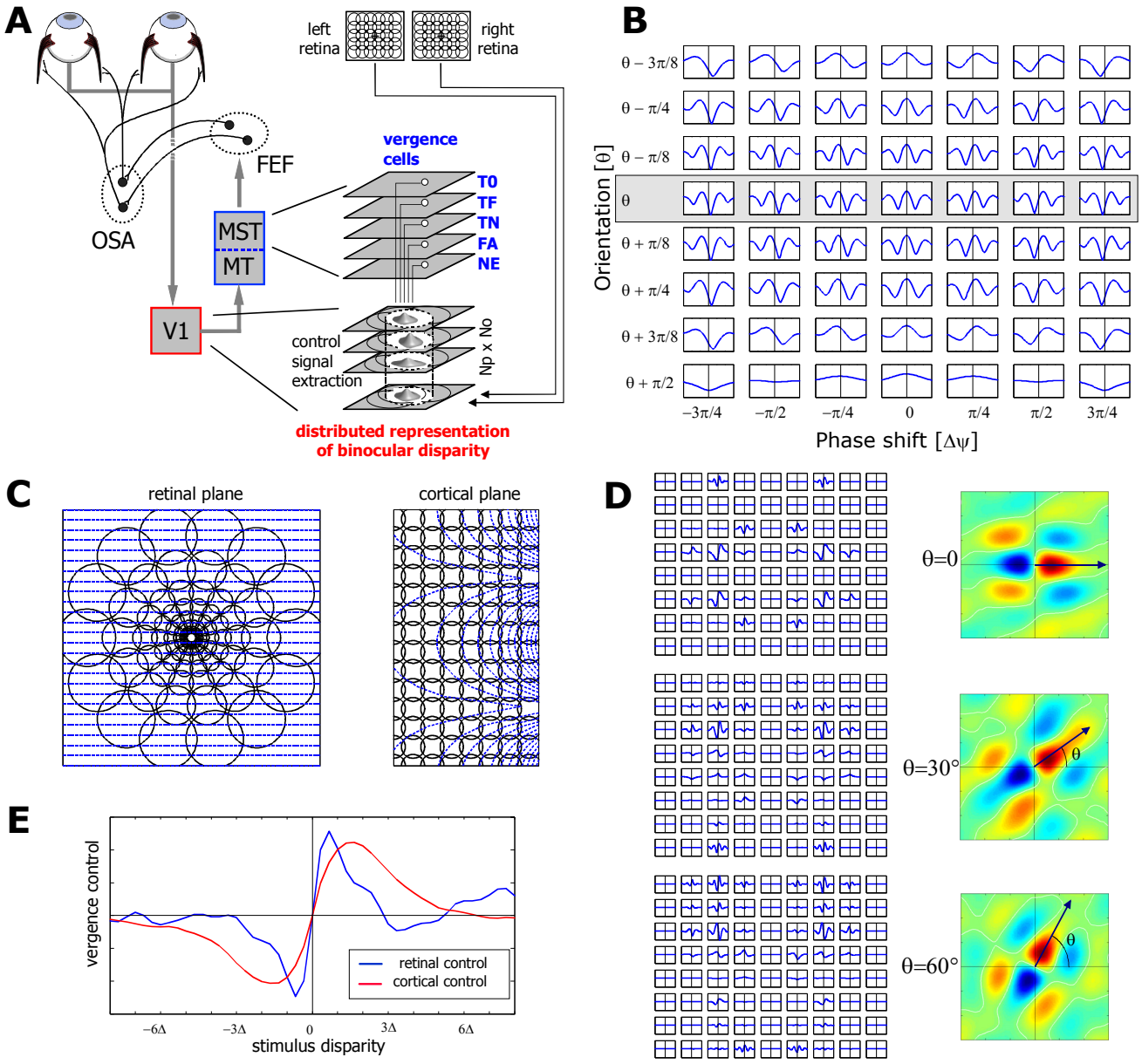


Figure 29: ***(A)*** *Schematic modeling of the neural circuitry involved in the control of vergence in primate brain.* ***(B)*** *Tuning curves of the population of binocular energy cells. The stimulus disparity is varied along the direction orthogonal to the orientation θ of the receptive field in the range [−3Δ, 3Δ], where ±Δ. is the maximum theoretical limit for disparity detection. The cell that are most modulated by the stimulus disparity, are those with the same orientation of the stimulus (gray area).* ***(C)*** *Receptive fields used in the model: a space variant sampling in the retinal domain with receptive fields of non-uniform size (left) corresponds to a uniform sampling in the cortical domain with constant size receptive fields (right).* ***(D)*** *Examples of different weightings of the population's tuning curves (left) depending on the desired orientation of the vergence control (right).* ***(E)*** *Comparison of the vergence control obtained by the same filters in the retinal domain (blue line), and in the cortical domain (red line).*

Starting from the model of a population of disparity detectors tuned to different orientations (see Fig. 29B), we can design an horizontal vergence control based on the detection of properly oriented disparities in the corical domain. More precisely, it is possible to design a different set of weight w_θ to get disparity-vergence curves properly steered along the direction of the mapped cortical disparity (see Fig. 29D, right) from a space-variant combination of the same set of basis disparity tuning functions used in the retinal domain (see Fig. 29D, left). Since the decoding in the cortical domain implies an increase of the receptive field size, the log-polar vergence control produces a correct behavior for a wider disparity range than the retinal one (see

Fig. 29E) with trajectories that resemble the psychophysical ones. In fact, it can produce a faster dynamic for large disparities, as they occur when the cells with larger receptive fields, centered in a perifoveal region are more stimulated. On the contrary for small disparities the model is able to provide a fine control and a stable fixation thanks to the smaller receptive fields, situated in a foveal region. Hence, the resulting control takes advantage of the elaboration in cortical domain since the compression of the information in the periphery reduces the computational cost, while leaving the precision of the control unaltered. Moreover, only the SHORT mode can be considered since the space-variant feature of the log-polar mapping guarantees a *structural* implementation of the dual-mode behavior.

Task 2.2: ***Interactive depth perception.***
The computer vision principles developed in the previous period have now been applied in a biologically plausible architecture. Retinal disparity is no longer explicitly calculated and the methods instead operate directly on the response of a population of binocular energy neurons. Image warping operations have been omitted as well. Using a learning approach, a feedforward neural network has been developed that can directly transform this population response, together with the gaze angles, into a 3D scene description based on head-centric disparity. Furthermore, the same architecture enables the extraction of (a limited set of) gaze angles directly from these responses. Since the mechatronic system is not available to us for demonstration purposes, we have applied the autocalibration algorithm to real-world image streams obtained from the iCub-platform (Nosengo, 2009). This same platform is also used to demonstrate the vergence mechanisms from Task 2.1. The methods proposed here can operate together with these vergence mechanisms in various ways. Improved calibration estimates can feed directly in the convolutional network for vergence control presented in Deliverable 2.1, but can also modulate the weights of the mechanism (also reported there) that integrates the population responses into the vergence control.

In the remainder we give a brief overview of the network architectures and learning procedure, together with some results. More details can be found in Deliverable D2.2b.

Network architectures and learning procedure

The transformation from retinal to head-centric disparity requires compensating for the transformations induced by the (3D) rotations of the left and right eye. Since we operate on the responses of a population of binocular energy neurons, this transformation needs to be performed together with correspondence estimation. We have decided not to perform these steps separately, but to rather directly modulate the responses of a population of binocular energy neurons so as to solve both problems at the same time. The complexity of this modulation warrants a learning approach.
Gain modulation using basis function networks (Pouget & Sejnowski, 1997) is not feasible here due to the large number of oculomotor signals that need to be combined with the population response. This leads to an explosion of dimensionality and a more efficient approach is required. We use a traditional black-box approach based on multi-layer perceptrons. Figure 30 provides an overview of the inputs and outputs and the network architecture used for head-centric disparity estimation. This network combines the population response (which implicitly codes for retinal disparity) with the oculomotor signals (the gaze angles) into the head-centric disparity. A similar network was used for gaze estimation, in which the oculomotor signal input was removed, and the output replaced by the gaze angles (see Deliverable D2.2b).
Due to the complexity of the transformation that needs to be learned, a large number of examples are required. It is therefore not feasible to use real-world images, and we instead generated synthetic random textures, warped in different ways, for training the network. These textures are first processed by a population of simple and complex cells. The population response is of high dimensionality (8112 dimensions), many of which are strongly correlated. We therefore applied principal components analysis (PCA) to reduce the dimensionality. The resulting components then serve as input to the network, together with the oculomotor signals.

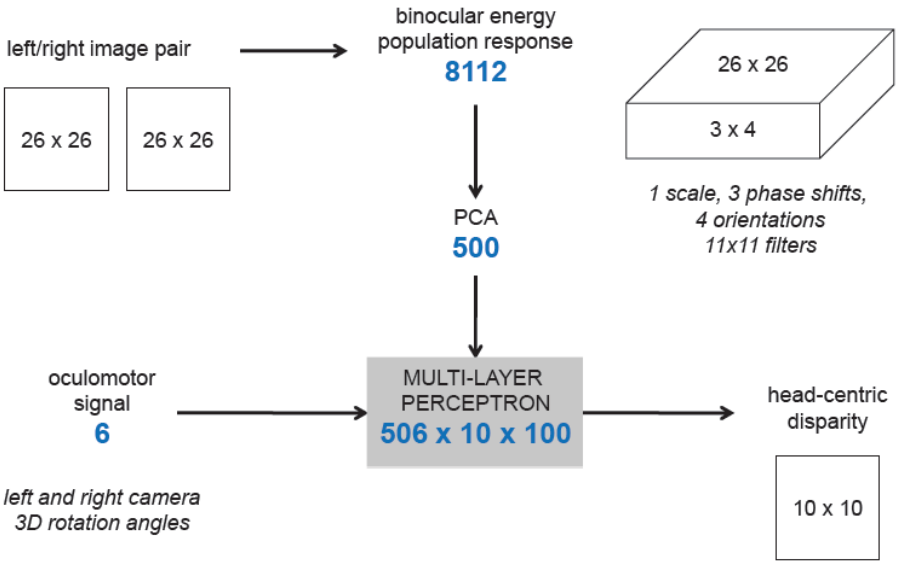


Figure 30: *Training procedure and network architecture employed for head-centric disparity estimation.*

Results: Head-centric Disparity Estimation

We use a total of 3750 samples for training the network, which is randomly divided in training (70%), validation (15%), and test set (15%). We then evaluate the performance on a completely independent test set consisting of 1250 samples. To demonstrate that the network is able to correctly apply gaze information, and to show the importance of this information, we compare the performance of the network with gaze information shown in Figure 30, to a network that only has access to the population response (after PCA). The head-centric disparity map estimated with both networks is shown in Fig. 31(B,C) together with the ground truth head-centric disparity (Fig. 31A) for five typical samples from the test set. Note that the network without gaze input is able to predict the general magnitude of the disparity field, but cannot estimate its fine structure. The network with gaze input performs the transformation with a much higher precision. On the complete test set, the correlation coefficient between the estimates and ground-truth is equal to 0.9192 without gaze input, and 0.9872 with gaze input.

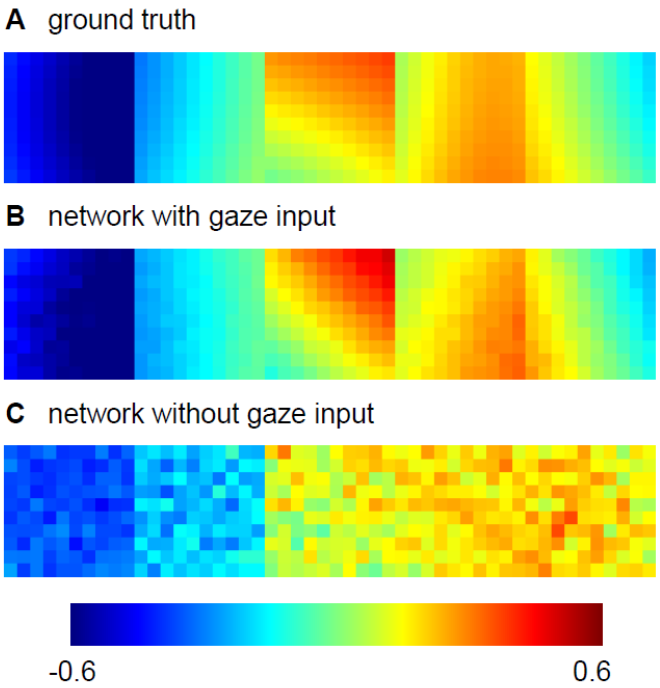


Figure 31: *Ground-truth head-centric disparity **(A)** and head-centric disparity predicted by a network with **(B)** and without **(C)** gaze input.*

Results: Gaze Estimation

For gaze estimation, we remove the gaze input from the network, and replace the target head-centric disparity with the (at most) six gaze angles. The gaze information estimated in this way from the population responses can be used to correct the imprecise information available through proprioceptive feedback from the motor system, which enables better vergence/version control and more precise correspondence estimation.

This problem is notoriously ambiguous, and only the essential matrix can be extracted from image data. Therefore, we have explored a set of problems with gradually increasing gaze complexity. The results are shown in Fig. 32. From this figure, we can see that the performance is quite good when only one eye is considered at a time. In the top row, a single gaze parameter is changed, while all the other remain zero. Both tilt (Fig. 32A) and torsional (Fig. 32C) rotations are easy to predict, because they introduce a strong vertical disparity pattern. Pan movements on the other hand are more difficult (Fig. 32B), because they are more easily confused with the disparity pattern. This however should not affect the precision of correspondence finding, and so is less relevant in this context. The network is also able to simultaneously estimate all gaze parameters of a single eye (Fig. 32D) with a performance similar to the worst single parameter performance. We also examined the degree to which all gaze angles for both eyes can be predicted together, but (as expected) this did not yield very good performance (Fig. 32E).

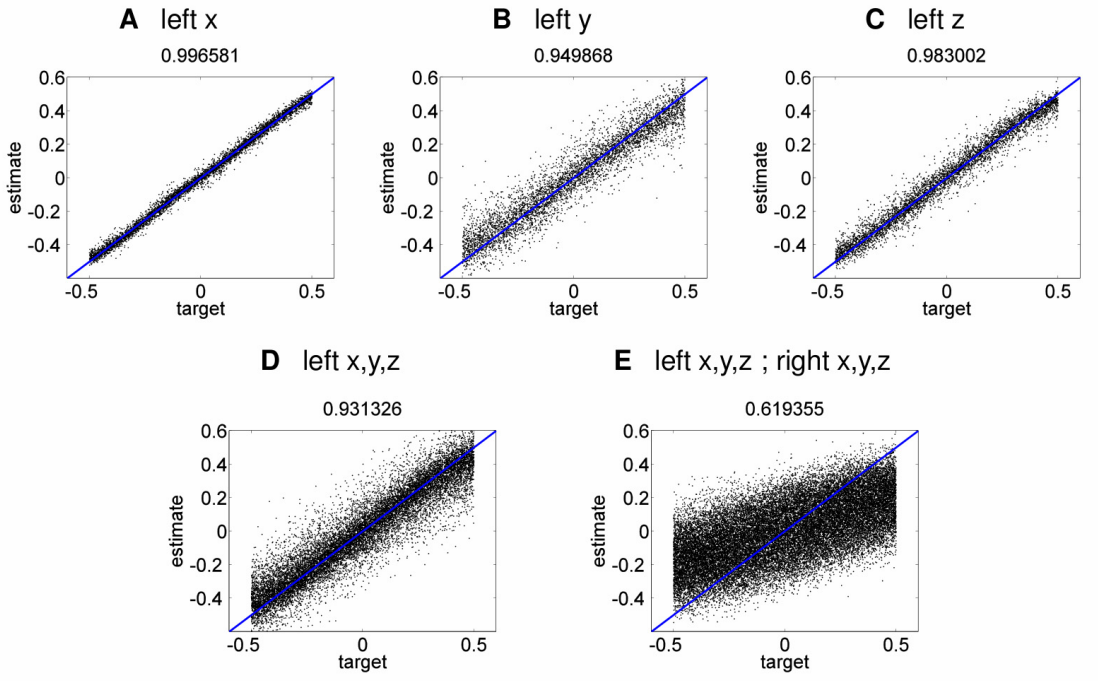


Figure 32: *Ground truth versus estimated gaze angle scatter plots for training scenarios of different complexity. In the top row, only a single gaze parameter of the left eye is changed: tilt (A), pan (B), or torsion (C). In the bottom row, all the left eye parameters are changed in (D), and both left and right gaze parameters are changed in (E). Only test set data is shown, and the correlation coefficient obtained on this set is indicated above each figure.*

The ability to estimate gaze in the presence of complex disparity patterns, appears to be quite feasible following this approach for each eye only. This means that also autocalibration is possible with this method, since in the computer vision approach we developed earlier, we used an alternating approach to correct the gaze estimate by considering one eye at a time.

Deviations from the project workprogramme

None.

WP3: Selecting and binding visual fragments

Leader: Fred Hamker (WWU)
Contributors and planned/actual effort (PMs) per participant:
WWU (17/17), K.U.Leuven (3/3) and UG (1/1)
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 us 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).

Progress towards objectives

Task 3.2: ***Selecting visual fragment***
This task has the objective to investigate and develop mechanisms of attentional selection. In the previous, second period, we had shown that a concept of feature-based attention (Hamker, 2005) can be combined with a learned object descriptor for attentional selection at the object level. In this proof of concept we were able to select objects among several distractors in a virtual reality setup. We extended this concept in the third period to attentive stereoscopic object recognition (for details see deliverable D3.2, Beuth et al., 2010 [C12]) and investigated the applicability of this approach on natural scenes. Thus, the model used in the VVCA (Vergence-Version Control with Attention effect) has been refined and ported to the robot setup of partner UJI. This section summarizes the system designed for the robot. Concerning technical issues, the porting of the system was straight forward from MATLAB to C++. The module communicates via the Yarp framework and requires less than one minute to process one image. Real time processing was not desired, therefore no optimizations in this regard has been taken into consideration.

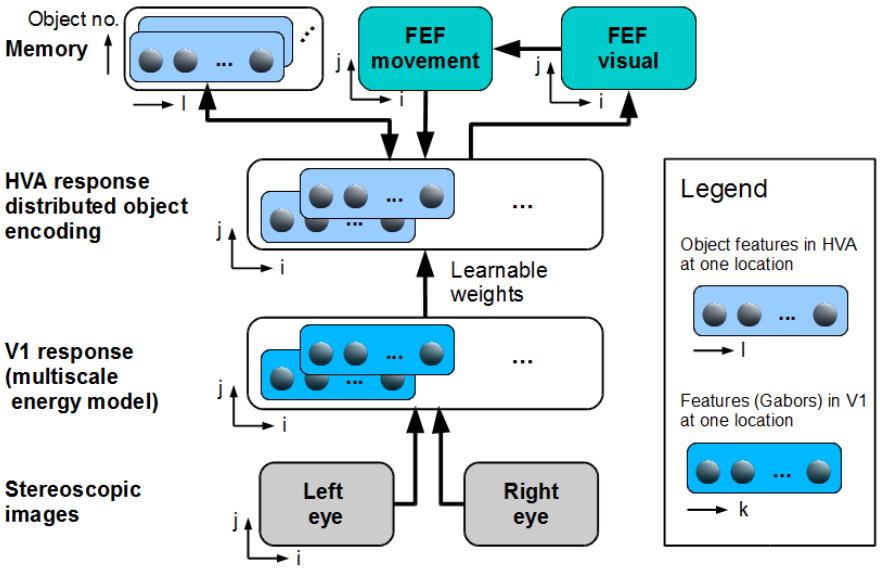


Figure 33: *Neuronal model of the stereoscopic object detection. The i and j indices correspond to the spatial x and y axis of the images. The index k refers to different Gabor responses and l to different learned features*

in HVA. Adapted from Beuth et al., 2010.

The model: the object recognition system (Fig. 33) uses learned object representations based on a V1 stereo energy model. This energy model (Chessa *et al*., 2009b; Sabatini *et al*., 2010) provides local information about disparity and frequences. Unlike as in deliverable D3.2, the energy model uses multiple-scale filters to deal with larger disparities and objects. The high level area (HVA) can be compared to brain areas such as V2/V4/IT, whose cells can be interpreted as representing multiple views of a single object. We achieved the object selectivity by learning the connections from the energy model to HVA with a biological motivated learning algorithm (Wiltschut & Hamker, 2009) and a trace rule using temporal continuity for the development of view-invariant representations of objects (like in Földiak, 1991; Rolls & Stringer, 2001; Wallis & Rolls, 1997). We used a weight sharing approach to analyze the whole visual scene in parallel, i.e. the detection of objects is independent of the location of the object in the visual scene. A top-down "attention signal", which is simply stored for each object in memory, can bias a particular object selection. 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. 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 HVA. 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 HVA. After successful recognition, the FEF movement map encodes the position of the searched object.

Results – We have tested the recognition for four objects (the three ones in Fig. 34 and the grasping unit of the robot). The grasper of the robot is visible if the robot points to an object. We learned the grasper as a normal object which allows the partner UJI to ignore the grasper if necessary. Furthermore, UJI needs the position of the grasper to train the visuo-motor coordination (see WP4). Now, they could replace their marker using the object recognition system to localize the grasper position.
We tested 25 scenes with those 4 objects using a task that requires the successful recognition of each object in every scene. 98% of the objects were correctly classified. Fig. 34 shows for one scene the maps after successful recognition and Fig. 35 shows the recognition process itself over time.

Discussion – A known limitation of the energy model is that it can encode large disparities only at a scale with a low image resolution. However, at such low resolutions the shape of an object disappears and an object is mostly only represented by a broad blob of activity. Therefore, we suggest to add filters with larger disparities to the model, to use a space variant model, or to use learned filters in the energy model. The limitation in disparities of the energy model is also the reason for 2% false recognitions.

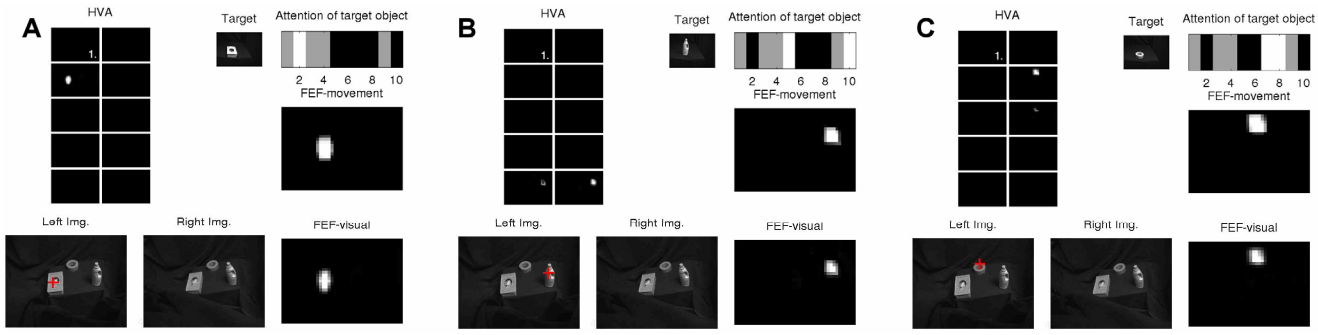


Figure 34:: *Successful localization for each of the 3 objects in a single scene. The figure shows the layer activities during the object localization experiment for the robot. The responses of the population code (10 cells encoding features) in HVA are computed as described in Fig.33. Thus each box shows the activity of a single feature in image coordinates. The graphs also show the attention signal (population code on the x-axis) and the two FEF maps. Normally, the x- and y-axis correspond to the spatial x and y axis of the images. A red cross marked a successful recognition and saccadic selection of the target object.*

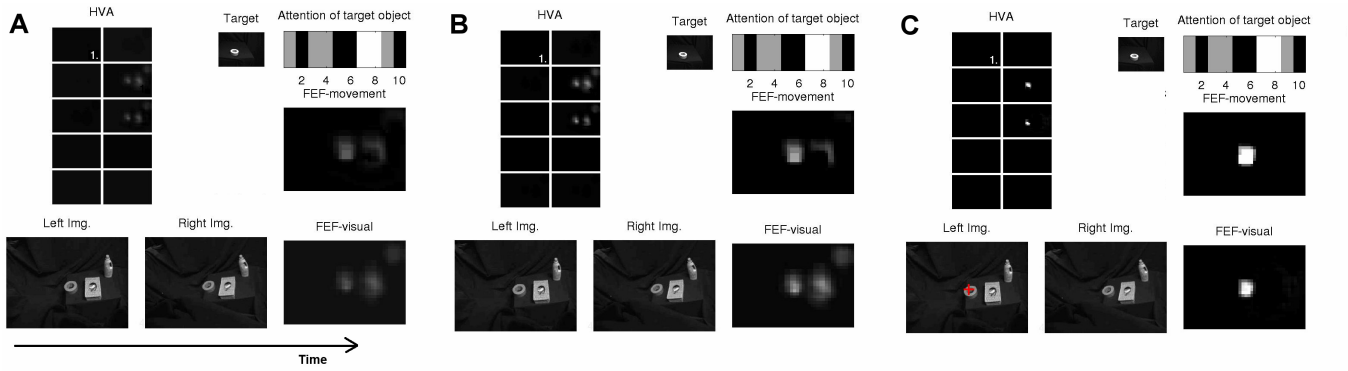


Figure 35: *Temporal dynamics of the model illustrating the spatial binding process and competition. For description see Fig. 34. Initally, some HVA cells representing a non-target object (the box) are activated as well. However, due to the attentional top-down signal, the target (adhesive tape roll) becomes more active over time and finally only the cells at one area (the location of the target) remain active and the cells at the other location were suppressed.*

The object recognition system uses only unsupervised learning (spatial and temporal statistics) for learning to distinguish the objects. Objects which do not differ in their shape or in their temporal occurrence cannot be learned by this model. In this case, an interactive exploring of the environment and/or reward-based learning is necessary, which has been investigated in Task 3.2. Moreover, learning should probably take place across multiple layers to obtain more invariances and still a robust recognition. Here, we have learned only within a single layer.

To summarize, we have shown how the attentive object recognition can be integrated into a robotic system and can select a particular visual object in a natural scene. The model combines stereoscopy, learned object descriptors and the concept of attention in a biologically plausible way. This still relatively simple system could be used in a wide range of applications in active vision.

Task 3.3: ***Selecting between behavioral alternatives.***
In second year of the project, we have developed a basic model (see deliverable D3.3a) to select between behavioral alternatives. This has been extended in the third year in two ways: 1) we have combined this working memory model (deliverable D3.3a, Vitay & Hamker, 2010 [J14]) with the object recognition system from deliverable D3.2 (Beuth *et al*., 2010 [C12]). 2) We have expanded it with a biologically grounded model of working memory (Schroll *et al.,* 2011 [C12]).

Combined model

The goal of the combined model (Fig. 36) is to maintain Working Memory (WM) and to select the correct response for the current task. The model learns from visual experience and from previous rewards by associating these rewards to specific visual stimuli. The central idea of maintenance of the working memory is that an agent has to remember objects that are useful or necessary for the task at hand. An object can be defined as very useful, if the agent can expect a high reward if it remembers this object and chooses a decision based on it. By this idea, the agent learns when it should store, hold or delete a specific stimulus from WM.
We have built objects in a virtual reality (VR), each object representing a symbol for the task. First, the stereoscopic views from the VR are processed by the above described vision model. After recognizing the object, we pass the activations to the WM and motor response selection model. The working memory is organized in several cortical-basal-ganglia-thalamo-cortical loops (WM loop) which learn to maintain a certain symbol. The model learns to store the symbols associated with high reward and suppress the ones with low reward from WM content. The exact functionality will be discussed in the section "Biological architecture...". From the current presented stimuli and the past sensations stored in WM, the model learns to associate a motor response. By learning, the model will choose behavioural responses associated with high reward.

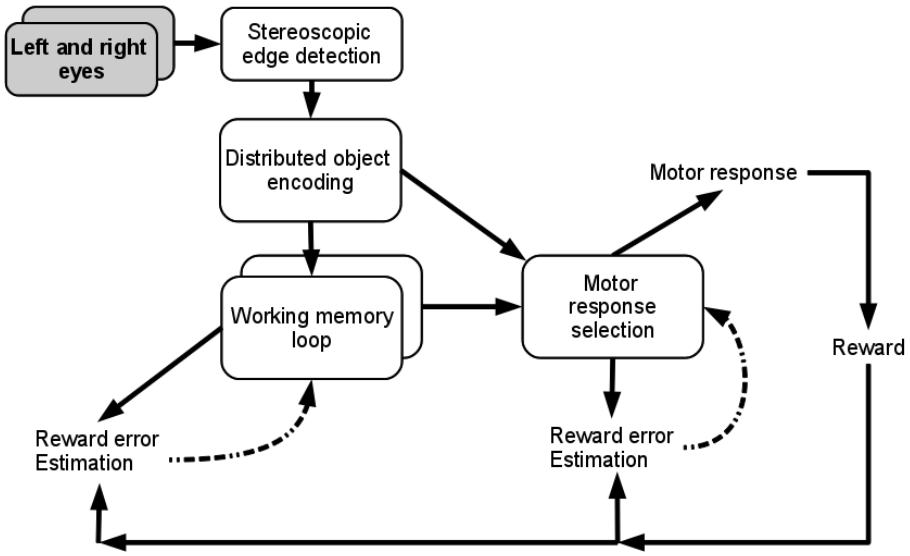


Figure 36: *Functional overview of the combined object recognition (stereoscopic edge detection and distributed object encoding) and working memory model (working memory loop and motor response selection). Dashed lines indicate Dopamine influence, therefore these weights are adapted according to the received reward.*

Biological architecture and function of the model

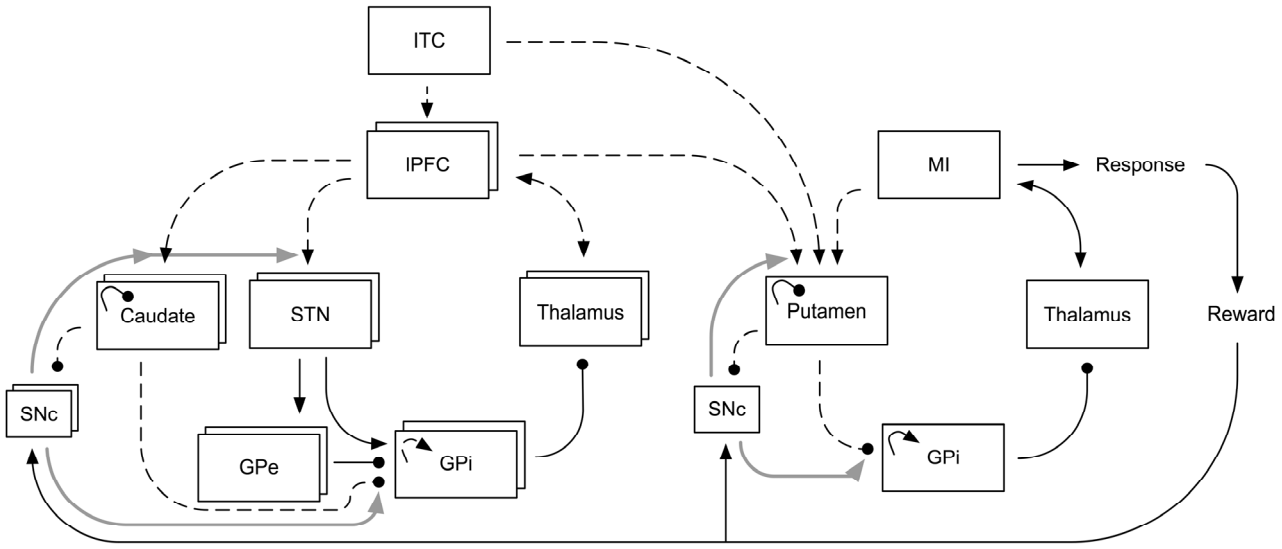


Figure 37: *The architecture of the working memory model consisting of prefrontal cortico-BG-thalamic loops (WM) and a motor loop (choosing between behavioural alternatives). Solid arrows denote hard-coded connections between or within layers, dashed arrows learnable ones. Pointed arrows symbolize excitatory connections, rounded arrows inhibitory ones. The gray arrows deriving from SNc represent the Dopamine influence on learning within BG afferents. GPe: globus pallidus external segment; GPi: globus pallidus internal segment; lPFC: lateral prefrontal cortex; MI: primary motor cortex; ITC: inferior temporal cortex; SNc: substantia nigra pars compacta; STN: subthalamic nucleus.*

Basal ganglia (BG) can be divided into multiple functional domains based on cortico-striatal afferents (Alexander *et al.,* 1986). For modelling a WM task, the major ones are 1) the executive domain (associated to the caudate nucleus as another part of striatum) is mainly connected to lPFC. It is involved in goal-directed learning, action-outcome associations and WM (Redgrave *et al*., 2010). 2) The sensorimotor domain is mainly associated to premotor and sensorimotor cortices and is involved in action selection and stimulus-response associations (Horvitz, 2009); the corresponding striatal part is called putamen. These functional

domains interact through ascending cortico-cortical projections, thalamo-cortico-thalamic projections and through a spiraling pattern of connections between striatum and the dopaminergic areas of SNc (Haber, 2003).
The model (Fig. 37) consists of two prefrontal cortico-BG-thalamic loops (PFC-loops) and one motor loop that have the same general architecture and obey the same learning rules. PFC loops (left of Fig. 37) control WM by flexibly switching between maintenance and updating of information. Then, they bias a motor loop (right of Fig. 37) to decide between a set of possible responses. The loops' functional architecture works as follows. Activation in the cortex excites striatal and subthalamic neurons. Striatum (Putamen and caudate nuc.) inhibits tonically active neurons of GPi (direct BG pathway) which in turn excites thalamic neurons. In contrast, activation of STN causes a fast excitation of GPi and a slower inhibition via GPe (hyperdirect pathway). Therefore, the hyperdirect pathway gives a brief and global reset pulse to GPi, where the direct pathway allowing to maintain symbols in the loop.
Most eminently for learning of behaviours, BG has an important role in reinforcement learning: BG receive dopaminergic afferents from SNc providing the BG with an error signal of reward prediction (Schultz *et al*., 1997, Hollerman & Schultz, 1998, Schultz, 2002). This reward prediction error is encoded by Dopamine levels which have been shown to modulate long-term synaptic plasticity within BG (Reynolds *et al*., 2001; Surmeier *et al*., 2007; Shen *et al*., 2008). Functional, Dopamine encodes the difference between the expected and the currently received reward. Dopamine bursts (above a tonic baseline level) occur from unexpected rewards while dopamine depletions (under this baseline) follow omissions of expected rewards. The WM model learns to maximize the received reward for a task by estimates the expected reward for each symbol. If the model receives more reward than expected, the model reinforces Dopamine-modulated connections (marked by grey arrows in Fig. 37) which in turn reinforces the memorization of a certain symbol in a certain loop. The idea is that the object was helpful to solve the current task and it will also be useful in the future, thus the model should remember it. On the opposite, if the model receives less reward than expected, the model has remembered the wrong object resulting in an inhibition of Dopamine modulated weights and the model will less likely store the symbol in working memory in the future. For details like learning rules, please refer to D3.3b or (Schroll *et al*., 2011 [J23]).

1-2-AX Task

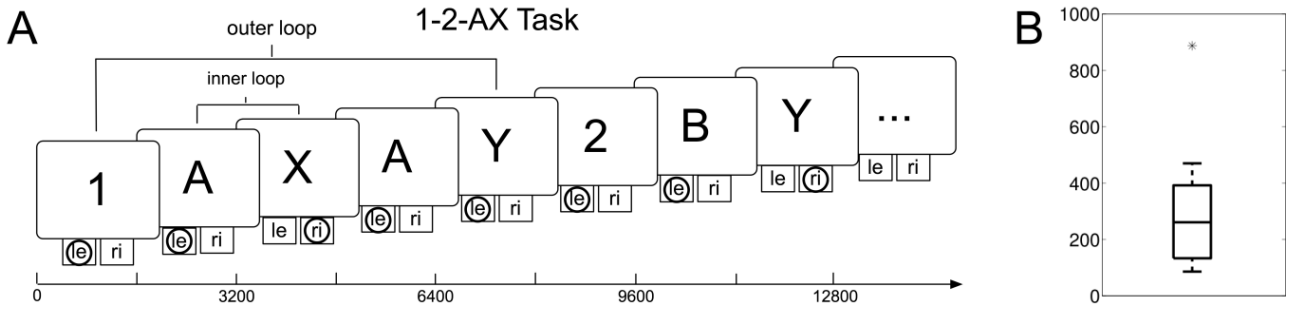


Figure 38: ***(A)*** *The 1-2-AX conditional WM task: in each trial, a stimulus is presented and the model has to choose between a left- and a right-button press. Circles indicate correct responses. A response of 'right' is only correct for the two combinations 1-A-X and 2-B-Y of the symbols, otherwise 'left' is correct.* ***(B)*** *The combined model's performance in learning of the WM task 1-2-AX. Box plots show the number of trials needed until the last error occurs. The boxes' upper and lower borders represent upper and lower quartiles, respectively. Outliers are represented by asterisks.*

Sharing a peripersonal workspace is one of the goals of the project EYESHOTS, which requires to hold previously visible information in memory to allow the agent to be able to choose between behavioural alternatives. Both requirements are addressed by the working memory model. To ensure that the task is general enough and also replicable, we decided to use a well known task from literature of working memory. In this 1-2-AX task (O'Reilly & Frank, 2006), see Fig. 38A, decisions must be taken dependently on the previously presented symbols and the agent must be able to deal with distractors (distractors are irrelevant objects for the current task which should not be remembered). Only special combinations of the symbols (e.g. a '1', followed by '2' and by 'X') require a particular response (here, press the right button), all other combinations result in the other behavioural alternative (left button). The number of possible combinations is

very high and the agent does not know in advance if a symbol is important or irrelevant. This is also typical for real world tasks which make such tasks very challenging. Figure 38B shows the performance of the combined model of 20 randomly initialized networks successfully learning the 1-2-AX task. Networks not learned the task to criterion were removed from the data which occurs for approximate 25% of the networks.

Summary

We have extended the approach presented in deliverable D3.3a by proposing a biologically computational model that combines object recognition and reinforcement learning. Our model demonstrates that both flexible control of WM and adaptive stimulus-response mappings can develop within parallel, hierarchically interconnected cortico-BG-thalamic loops. Based on Hebbian- and Dopamine-like learning, prefrontal loops learn to flexibly control WM while a motor loop learns to decide between a set of possible responses. By teaching the 1-2-AX problem to the model, we have shown a formal task comparable to those that can occur in a shared workspace scenario.

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:
UJI (17.27/18.86), UNIBO (2/2.43) and UG (6/5.87)
Planned/actual Starting date: Month 1/1

Workpackage objectives

During the third year, Task 4.2 was completed, achieving the goal of generating 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. We fully implemented in 3D the visual/oculomotor and oculomotor/arm-motor basis function networks which allow bidirectional transformations between retinotopic, head-centered and arm-centered reference frames. We ported the networks on the robot platform, enabling the UJI humanoid robot *Tombatossals* to accurately learn the transformations between visual, oculomotor and arm joint spaces by actively interacting with its surrounding environment. Adapting the architecture and parameters of the networks to the findings of WP5 regarding V6A and the coding of space (to which UJI directly participated, Bosco et al., 2010), we were able to reproduce some psychophysiological effects, such as those related to saccadic adaptation experiments, described with more detailed here below.

Task 4.3 extends the skills of Task 4.2 to the exploration of visual stimuli in the surrounding space. The agent simultaneously learns to reach towards different visual targets, achieving binding capabilities through active exploration, and builds an egocentric "visuomotor map" of the environment. A tighter interaction with the partners in charge of developing WP1, WP2 and WP3, achieved also thanks to the advice of the reviewers, permitted us to implement more advanced visual and visuomotor skills that make the robot able to interact with multiple real objects in a simplified working-desk setup. The global abilities of the robot at the end of the project, according to the goals of Task 4.3, will be presented in a live demonstration during the final review meeting, as described in detail in deliverables D4.3b and D4.3c.

Progress towards objectives

Task 4.2: ***Generating visuo-motor descriptors of reachable objects***

A fundamental goal of Workpart 4 was to design and develop a model for representing the robot own movements and the nature of the surrounding environment by using eye and arm movements. The model, described in detail in deliverables D4.2 and D4.3a, and in Chinellato *et al.* (2011b) [J3], was fully implemented on the "Tombatossals"' humanoid robot setup (Antonelli *et al.,* 2011 [C9]).

Basic skills such as concurrent or decoupled gazing and reaching movements toward visual stimuli have been acquired by the robot. As planned, the robot is able to learn a visuomotor representation of nearby objects, and shows its capabilities by performing oculomotor actions toward visual targets placed in its peripersonal space, or toward the location where its hand lies. Moreover, it is also able to perform arm reaching movements to visible objects, either with or without gazing at them. This ability is achieved through a purposeful exploration of the environment, which allows the robot to build a visuomotor memory of surrounding objects, described in Task 4.3.

The robot skills are attained by exploiting its sensorimotor coordination ability, provided by the transformations Visual⇨Oculomotor (V⇨O) and Oculomotor ⇔Arm motor (O⇔A), implemented within a radial basis function framework. To better highlight the system capabilities and their relation to the neural model based on neuroscience data and insights, two different experimental setups have been devised.

The first scenario is a *working-desk setup*, with simple objects, and full 3D movements for both eyes and arm. This setup allows the robot to show its fundamental visuomotor skills in its interaction with visible and reachable objects. In any case, to be able to perform a complete visual exploration of the environment, the robot has to pass first through a learning stage designed to develop its own sensorimotor coordination. This is done by training the visual-oculomotor and the oculomotor-arm motor neural networks. It is important to remind that, to reduce the learning stage at a reasonable amount of steps, the networks are bootstrapped with the weights obtained by the model of the system trained in a similar fashion. Also, although the networks work and the robot is able to act with good precision in all the visible and reachable space, it is convenient to

perform a specific traning session of some dozen trials focused on the fraction of space the robot is most likely to work within. A typical result of training of the different transformations in 3D is provided in Table 2, for a working desk-setup similar to the one of Fig. 11. In this case, the worsened performance of the Oculomotor⇨Arm motor transformation is rather apparent. This is very likely due to an unbalanced overlapping of the RBF centers, traning and test ranges for that particular working configuration. It is worth reminding though that the displayed results are achieved only with 7x7x7 neural networks. Still, as a future work, we plan to adapt dynamically the centers of the RBFs in order to better adapt the system to different working conditions.

Table 2: *Mean and standard deviation of the final error of all sensorimotor transformations after training with the robot.*

Transformation	Error (mean) [deg]	Error (std) [deg]
V⇨O	0.42	0.34
O⇨A	3.95	3.55
A⇨O	0.44	0.31

The second setup is aimed at showing in more details the properties of the radial basis function model applied to the real robot. This scenario sees the robot interacting with a computer screen similarly to what is done in typical psychophysical experiments. More precisely, we employed this setup to emulate the saccadic adaptation paradigms studied by partner WWU. The analysis of the robot behavior constitutes in fact also a contribution to cognitive science research, as demonstrated by the experiments presented below.

Saccadic adaptation experiments

In order to check the underlying properties of the computational framework on which the robot behavioral abilities are built upon, a different experimental setup was established. We opted for a cognitive science setup similar to those used for the saccadic adaptation experiments performed by partner WWU (Collins *et al.,* 2007, Schnier *et al.,* 2010). This consists of a computer screen placed within reaching distance, on which different visual stimuli associated to action signals are visualized. It is worthwhile to clarify that, although vergence varies just slightly in this setup, all the transformations are fully three-dimensional, and the robot keeps acting as in the usual 3D configuration. The comparison between data available from the human subjects, and those obtained by the computational simulation and from the robot can provide theoretical insights on aspects of the visuomotor cognitive behavior, and contextually allows us to validate our approach.

The exact experimental setup for simulating human saccadic adaptation experiments with the robot is the following. A computer monitor (1440×900, 19") is put in front of the robot at a distance of 720 cm, which allows us to obtain version angles similar to those occurring in human experiments without getting too close to the image periphery. The experiment program was designed to display at required positions small red squares (5×5 pixels), unambiguously identified by the robot *blob detector* module.

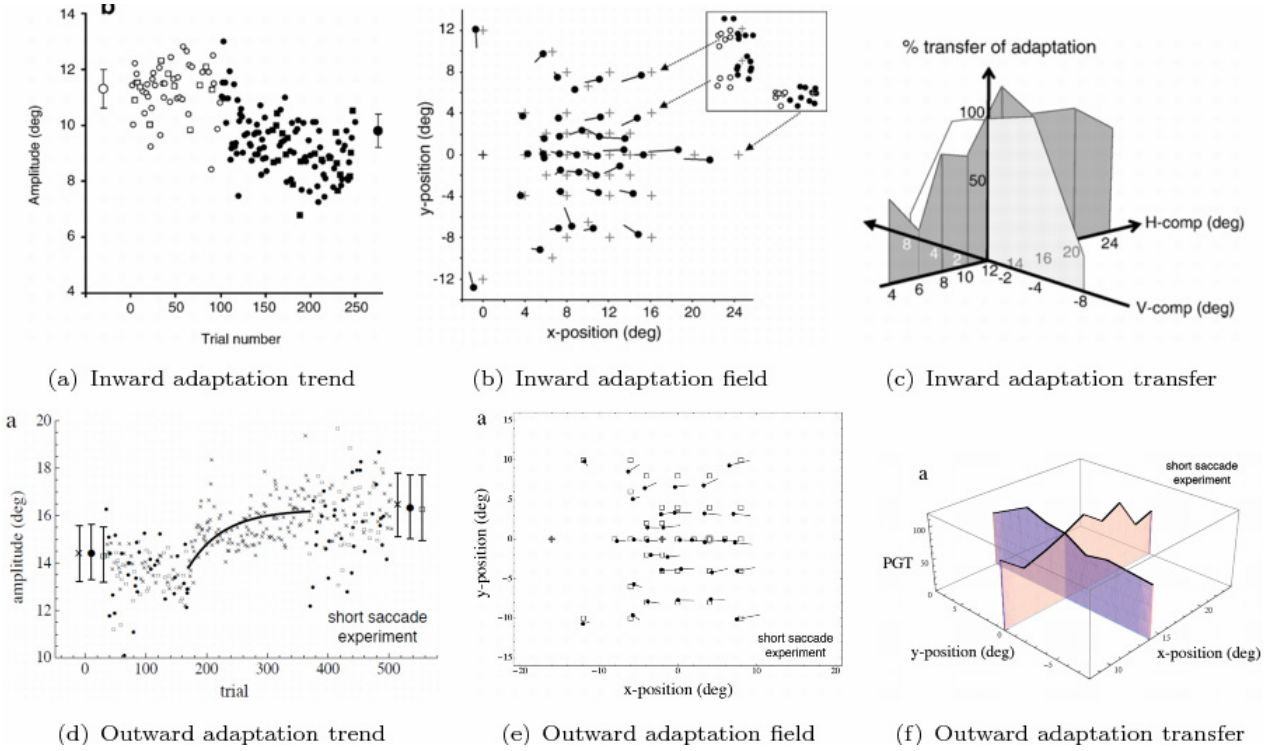
The task starts with the robot fixating a starting point stimulus (*FP*), placed exactly in front of the robot, so that it corresponds to a null version angle. A second visual stimulus is then displayed at the target position (*TP*), having the same vertical coordinate, but a displaced horizontal position by a certain amount x_1, which is a parameter of the experiment. The robot is required to perform a saccade toward this new stimulus. When the saccade movement signal is released, the stimulus is displaced toward a third point (*DP*), either closer or further on the *x* axis with respect to *TP* (for inward and outward saccadic adaptation protocols, respectively). At the end of the saccadic movement, the robot thus perceives a visual error between its final position and the visual target, that should be at visual coordinates (0,0) if the saccade were correctly executed. Such residual visual difference is used to adapt the weights of the network performing the transformation from retinal to oculomotor coordinates. The starting stimulus is then displayed again and the robot saccades back toward it. The whole sequence is repeated 100 times.

Simulated experiments

Before performing the saccadic adaptation tests with the real robot, we simulated them using the robot model on a corresponding virtual setup. This simulation is useful for predicting the sort of experimental results that the real robot are expected to provide. On the one hand, this is done to avoid keeping the robot busy with the execution of irrelevant experiments. On the other hand, it allows us to assess the impact that real world tests

have on the purely theoretical insights provided by the simulation. For this reason, we offer a comparison between data from human experiments, from the simulation and from robot tests.

The reference graphs for human experiments are reproduced in Fig.1, above for the inward protocol (Collins *et al.,* 2007), below for the outward protocol (Schnier *et al.,* 2010). For both cases three fundamental aspects are analyzed, the same that we will explore in our experiments. Fig. 39(a,d) shows the adaptation trend, i.e. the time course of the gradual shift of the subject response from the initial movement amplitude to the displaced target, for inward and outward adaptation respectively. Saccadic adaptation fields, displayed in Fig. 39(b,e), assesses how the mis-trained movement affects saccades directed towards different targets in space. Black dots represent movement average endpoint before adaptation, whereas the end of the segments represents the average endpoint after adaptation. Finally, adaptation transfer is visualized in Fig. 39(c,f), where the amount of adaptation is shown as a function of the *x* and *y* coordinates of the saccade amplitude.



(a) Inward adaptation trend (b) Inward adaptation field (c) Inward adaptation transfer

(d) Outward adaptation trend (e) Outward adaptation field (f) Outward adaptation transfer

Figure 39: *Saccadic adaptation; human results for inward (above, from Collins et al., 2007) and outward adaptation experiments (below, from Schnier et al., 2010).*

We tested each of the two experimental protocols, inward and outward adaptation, with two different configurations of the visual to oculomotor radial basis function network. The *uniform* configuration has the centers of the basis functions distributed evenly on the input space (*x* and *y* cyclopean coordinates and horizontal disparity). In the *logarithmic* distribution, neurons are placed closer to each other at the center of the visual field and for small disparities. While for the uniform distribution all neurons have the same spread, in the logarithmic case radii vary according to the distance of a neuron from its neighbors. For what concerns the parameters of the experimental setup, the target was fixed for all experiments at 11.89° on the right of the starting point, while the displaced point was set at 7.96° for inward displacements and at 15.73° for outward displacements. These values were set considering the robotic setup, in order to generate eye version movements comparable to those measured in psychophysical experiments with humans. The learning rate α=0.001 was also the same in simulation as in real experiments. The results we obtained with our simulation are shown in Fig. 40. Adaptation trend, adaptation field and adaptation transfer (columns) are depicted for: uniform inward, uniform outward, logarithmic inward and logarithmic outward tests (rows).

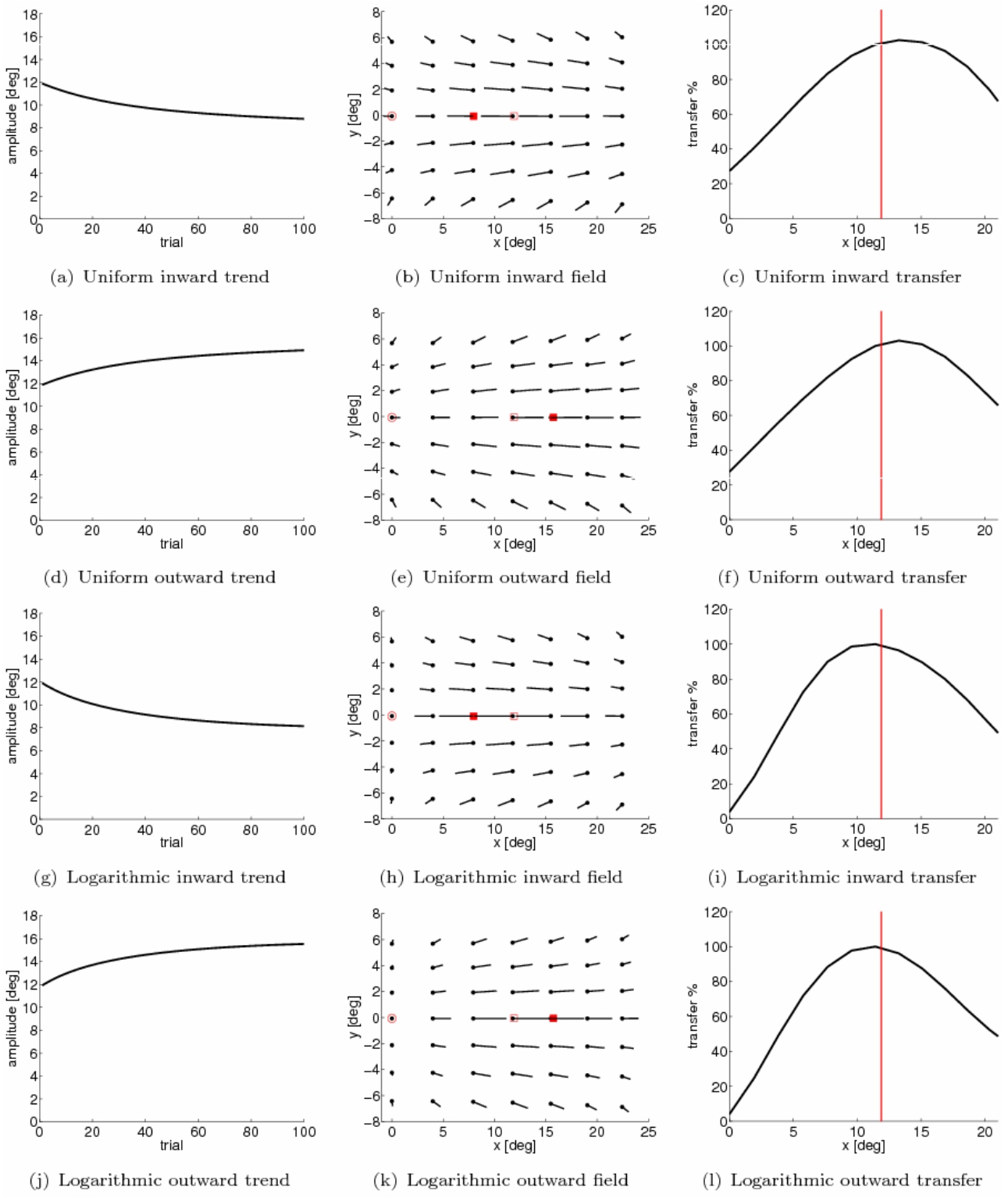


Figure 40: *Saccadic adaptation; simulated results with model.*

Adaptation trend graphs (Fig. 40(a,d,g,j)) show a plausible learning curve that reduces (in the inward case) or increases (in the outward case) the movement amplitude according to the deceiving feedback provided by the displaced target stimulus. In the uniform distribution tests, movement amplitude at trial 100 reaches to 8.79° in the inward protocol and to 14.92° in the outward protocol, from the initial 11.89°, for a final adaptation of 79.5% and 77%, respectively. The average adaptation over all trials is of 2.1° in both cases, about 54% of the target step. Slightly higher values (faster adaptation) have been obtained in the logarithmic case. In general, average adaptation values for humans are smaller than what we found in our simulations.

For the inward case, only a 13% adaptation was observed (Collins *et al.,* 2007), whilst 33-45% adaptations were registered for outward experiments (Schnier *et al.* 2010), depending on the initial saccade amplitude.

Adaptation fields (Fig. 40(b,e,h,k)) have in all cases a perceptible radial trend, with a *y* component indicating wider movements toward the top or the bottom of the screen for both protocols and net configurations. This is rather consistent with Schnier and colleagues outward tests (see Fig. 39(e)), but much less apparent for what concerns inward experiments (Fig. 39(b)).

The overall trend of the adaptation over the horizontal (*x*) component can be observed in the adaptation transfer graphs of Fig. 40(c,f,i,l). Human experiments suggest that, both for inward and outward adaptation (see Fig. 39(c,f)), the differences between pre- and post-adaptation movements peaks just after the abscissa of the target. Also, while transfer decreases with the distance from the peak, such decrease is slower for larger saccades than for shorter ones (gentler slopes on the right side of the peak). It seems that the uniform configuration captures the first of this phenomena, showing a peak in transfer for movement amplitudes slightly further than the target. Still, the transfer looks symmetrical with respect to the peak. The opposite occurs for the logarithmic distribution, which transfer peak appears slightly before the target abscissa. The transfer trend is though asymmetrical, showing a less pronounced decrease on the right of the peak. As observed in both inward and outward studies on humans, the adaptation vertical (*y*) component had a very small error rather homogeneous for different movement amplitudes, with no clear trend worth visualization.

Robot experiments

The same two configurations of the visual to oculomotor Radial Basis Function Network (RBFN) employed in the simulation were used also in the real robot experiments. The adopted configurations were chosen through a thorough search of center location and spread, because of their high precision in approximating the goal function. Their parameters are reported in Table 3.

Table 3: *Ranges and radii of the RBFN for the visual to oculomotor transformation in the saccadic adaptation task.*

Component	Range[*px*]	No. of Centers	Radii Unif.[*px*]	Radii Log. (max)[*px*]
Cyclopean X	±500	11	520	510
Cyclopean Y	±200	7	270	320
Disparity	±100	7	160	160

The network weights found on the model, and used in the simulated saccadic adaptation experiments described above, were transferred to the real robot. A short training phase with on-screen visual stimuli was then executed in order to adapt the network to possible distortions and unavoidable differences between the model and the real robot setup. This was performed by randomly showing a sequence of points on the screen, which the robot had to saccade to. The possible residual error after each movement was employed to train the network.

As in the simulation, four saccadic adaptation experiments were conducted with the robot, characterized by the basic structure of the visual to oculomotor network (uniform or logarithmic distribution of the centers) and by the direction of the displacement (inward or outward). Target and displaced point were the same as above: initial target 11.89°, inward displaced point 7.96°, outward displaced point 15.73°. Again, all experiments were performed with the learning rate set to α=0.001.

All results are depicted in Fig. 41, where the rows and the columns match the correspondent graphs of Fig. 40, obtained in the simulation for the same conditions.

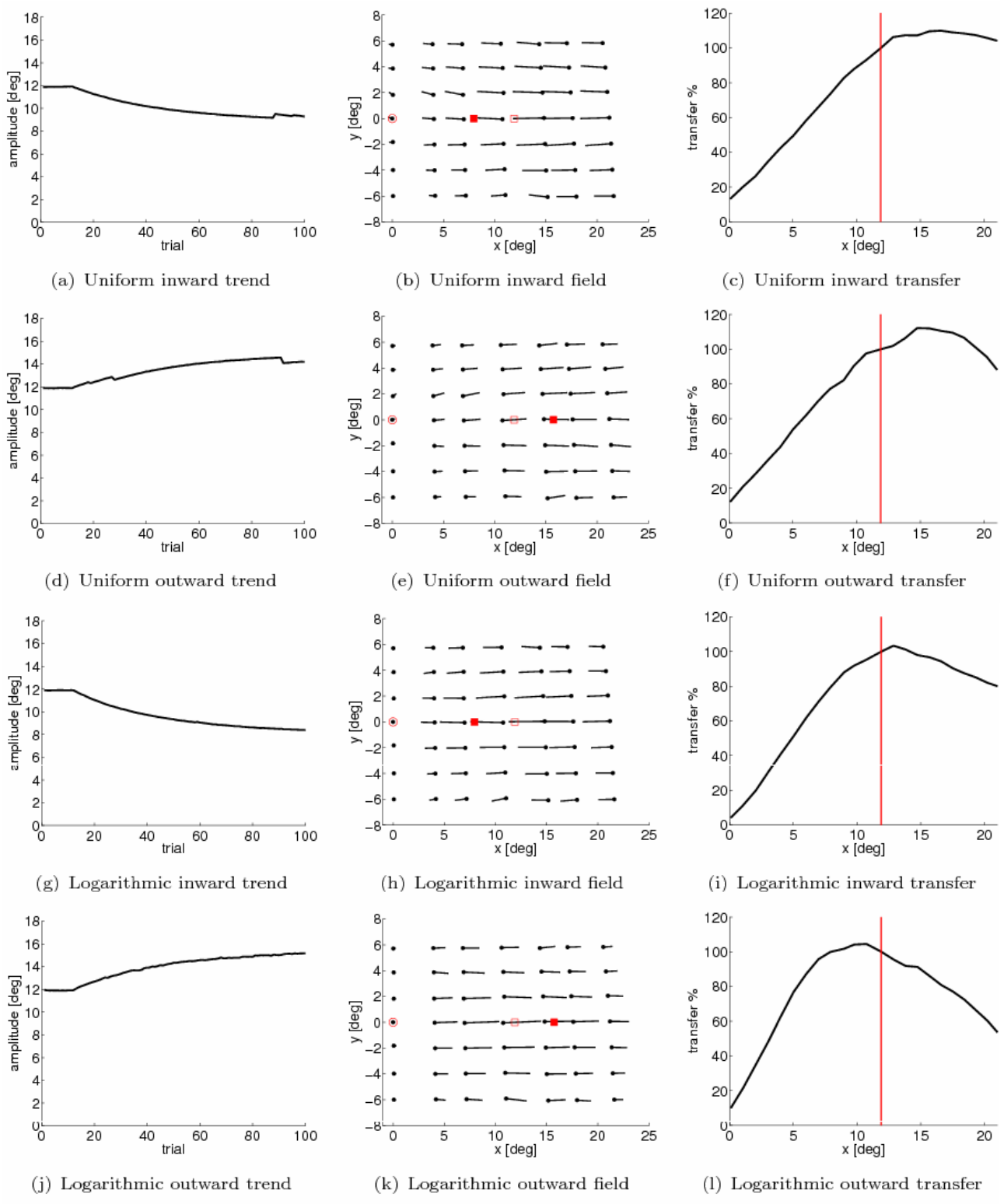


Figure 41:. *Saccadic adaptation; experimental results with the robot.*

The adaptation trend, shown in Fig. 41(a,d,g,j) for the four different tests, is very similar to what observed for the model and in human experiments. The final and average movement amplitudes are 66.5% and 43.0% for inward and 60.5% and 40.7% for outward adaptation, respectively. These smaller values suggest that, employing the same learning rate, the robot achieves a better approximation of the human data with respect to the simulation. Again, the logarithmic network provides higher adaptation values.

To study the adaptation transfer, an adaptation field was created by defining a 20x25 lattice on the screen, in order to evaluate the effects of adaptation on different potential fixation points. Starting from *FP*, all points on the lattice were shown one at time, and the robot was required to perform a saccade toward each stimulus.

At the end of the movement, the visual position of the stimulus and the oculomotor angles were compared. This process was performed before and after saccadic adaptation, but could be executed at any one of the 100 steps of the experiment, in order to monitor the progress of adaptation transfer. It is important to clarify that, during this evaluation task, learning is suspended and the network is frozen in its current state. This solution allows us to monitor precisely the evolution of the saccadic adaptation learning process, and constitutes thus an advantage with respect to human experiments, where such freezing is clearly not possible. Adaptation field are shown in Fig. 41(b,e,h,k), in which, for clarity reasons, only a subset of the lattice points have been visualized. The radial effect, when present, is very light and not consistent across different position, showing a pattern more similar to the human data than to the simulation results.

More interesting insights can be drawn by observing the horizontal (x) component of the movement change (Fig. 41(c,f,i,l). A late peak can be observed in both experiments for the uniform configuration (more pronounced than in the simulation) and also for the logarithmic distribution in the inward adaptation test. Moreover, practically all the cases exhibit an asymmetry of the transfer with curves descending more slowly for larger saccades, as in the human case. This effect is again stronger for the logarithmic network configuration. Once more, no relevant effects were observed for what concerns the vertical adaptation transfer component.
As a general consideration, it can be observed that the robot results approximate the human data better than the simulated results. The reduced radial aspect of the adaptation field and the trend of the horizontal component in peak position and slope asymmetry are more consistent between human and robot than the correspondent simulated results. This is especially interesting considering that exactly the same parameters were employed in the two cases. This phenomenon might reflect implicit properties of the hardware that affect the way untrained movements are biased by learning processes applied to similar movements.

Summarizing, different properties observed in the saccadic adaptation studies in human were captured by our tests. Both the simulation and the robot experiments showed plausible adaptation trends, slightly radial adaptation fields and typical features of the adaptation transfer on the horizontal component, such as asymmetry and late peak.

Task 4.3: ***Constructing a global awareness of the peripersonal space.***
In order to allow the UJI humanoid robot *Tombatossals* to perform the relatively complex behaviors required by the goals of Task 4.3, we have defined a global software architecture that permits the integration of modules of different levels of complexity, either internal or developed by other partners. The general working framework has been depicted in Figs. 9 and 10, and has been implemented in Yarp, as described in Section 3. Apart for the aspects already described above, we managed the allocation of attention and implemented a visuomotor memory in order to perform both gazing and reaching actions and also object recognition by integrating space (dorsal) with identity (ventral) visual information.
In fact, in addition to building a sensorimotor map of visual and motor targets in the nearby space, the construction of an integrated knowledge of the environment requires the identification of objects or targets and the use of memory of previously observed/reached objects. The behavioral schema in which object identity is associated to a particular sensorimotor configuration starts when the system is already able to accurately gaze, and possibly reach, toward a visual target. In this associative learning schema the robot is gazing at a given fixation point while two or more stimuli appear in its field of view. The robot is thus required to use its visuomotor skills, embedded in the neural networks that transform between visual and motor parameters, to estimate the eye motor movements required to fixate on each visible object, without actually executing the saccadic movement on the targets. The sums of the movement vectors with the actual gazing direction constitute an instance of the absolute positions of the visible objects. Next, the fixation point is changed; the robot gazes at the new one and estimates the new movements vectors required to fixate on the visual targets, creating new instances of the targets absolute position. The process is then repeated and at each step a slightly different absolute position is computed for each visual target. Ideally, all computed positions are the same, but due to unavoidable distortions and imprecisions we expect a range of variability, and the average of each estimated location is stored as the memory of the visual target position, for all targets. We employ a Kalman filter in order to maintain an average position that takes into account a certain number of instances while giving more credit to recent data. After the learning process has reached a minimum reliability threshold (usually in just a few steps), the system can be required to fixate and/or reach a target given its identity, using the visuomotor associations it has learned during the previous step. This can be done even on objects placed out of the field of view, if their location has been previously observed.

Deliverable D4.3b (Visual exploration of the environment) provides a description of the practical steps that compose this process of dorsal/ventral integration and creation of spatial awareness.

The above behavior is available to the robot thanks to the integration of the biologically-inspired modules (on stereo vision, vergence control, and object recognition) of partners UG, K.U.Leuven and WWU in the UJI framework. Comprehensive experiments in which the robot operates in multi-object setups creating its own visuomotor awareness of the environment have been executed and will be demonstrated live during the final review meeting. We will show how *Tombatossals* employs the egocentric representation of peripersonal space it has gained, to interact with surrounding objects, recognize them and perform custom visuomotor and arm-motor actions, such as: foveate on the hand; reach the gazing point; show memory; foveate on a given object (either inside or outside the field of view); reach a given object (either foveated or not); execute a sequence of saccades by employing either covert or overt attention.

Deviations from the project workprogramme

None.

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

Leader: Patrizia Fattori (UNIBO)
Contributors and planned/actual effort (PMs) per participant:
UG (0.5/0.5), WWU (13/13), UNIBO (6/6.7) and 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 for the design of artificial intelligent systems able to explore and interact with the 3D world.

Progress towards objectives

Neurophysiological experiments:
We proceeded with single cell recordings from medial parieto-occipital cortex (area V6A). We found that neurons there encode the 3D space through a variety of information: visual cues, attentional cues, oculomotor cues, and active arm movements performed in depth. Based on these findings, we suggested that V6A is able to coordinate eye- and arm-actions in the 3D space as well as to link, in a more general way, perception to action.

Psychophysical experiments:
We finished the investigation of the interconnection of visual fragments and motor parameter adjustment. 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 extended the investigation of the multisensory representation by investigating the reference frame for spatial eye movement control by demonstrating eye position effects in saccadic adaptation in both humans and monkeys. Or results show that saccadic adaptation takes place in supra-retinal references frames.
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. We collected and analyzed data both in a single actor setting and on human-human interaction. In addition we developed an experimental test for the social acceptability of a robot as an interaction partner via the social Simon effect.

Task 5.1: ***Role of visual and oculomotor cues in the perception of 3D space.***

1) Covert attention: a link between fragments without any effector movement

Attention is important for providing the link across single visual fragments, attention is used to select targets in a visual scene for prioritized processing and for preparing appropriately directed actions. In the third year of EYESHOTS, we measured the influence of covert attention toward different parts of the visual world in area V6A of the medial parieto-occipital cortex. We induced in the monkey covert shifts of attention in absence of any effector movement, neither of the eyes nor of the arm. We performed single cell recording in V6A, while controlling the monkey focus of attention addressing it toward several positions in the workspace. In this way, we found that neurons in V6A are influenced by the spatially directed attention (Galletti *et al*., 2010 [J10]). Figure 42 shows an example of a V6A neurons spatially modulated by the spatial shifts of attention toward peripheral locations, without concurrent shifts of the direction of gaze, a factor known to be powerful in modulating neural discharges in the medial parieto-occipital cortex (Galletti *et al*., 1995). Figure 42 shows a cell with a typical outward attention response for cues presented in the lower space. The spatially-tuned outward attention activity had a very long latency, and in some trials, the response

lasted until target onset, that is 1 s or more later than the cue onset. Although we cannot rule out completely that what we call outward attention response was a visual response to the cue enhanced by attention, the observed discharge was very different from a typical V6A visual response.
A neuron like the one reported in figure 1 may be particularly useful in the acquisition of peripheral visual information without shifting the gaze for the purpose of directing toward that location hand actions. This interpretatin is supported by recent findings on the involvement of area V6A in directing the hand toward targets located in different spatial locations (Fattori *et al.,* 2005), especially to non-foveated targets (Marzocchi *et al.,* 2008), in orienting the wrist and shaping the fingers to perform the appropriate grasp (Fattori *et al.,* 2009; 2010).

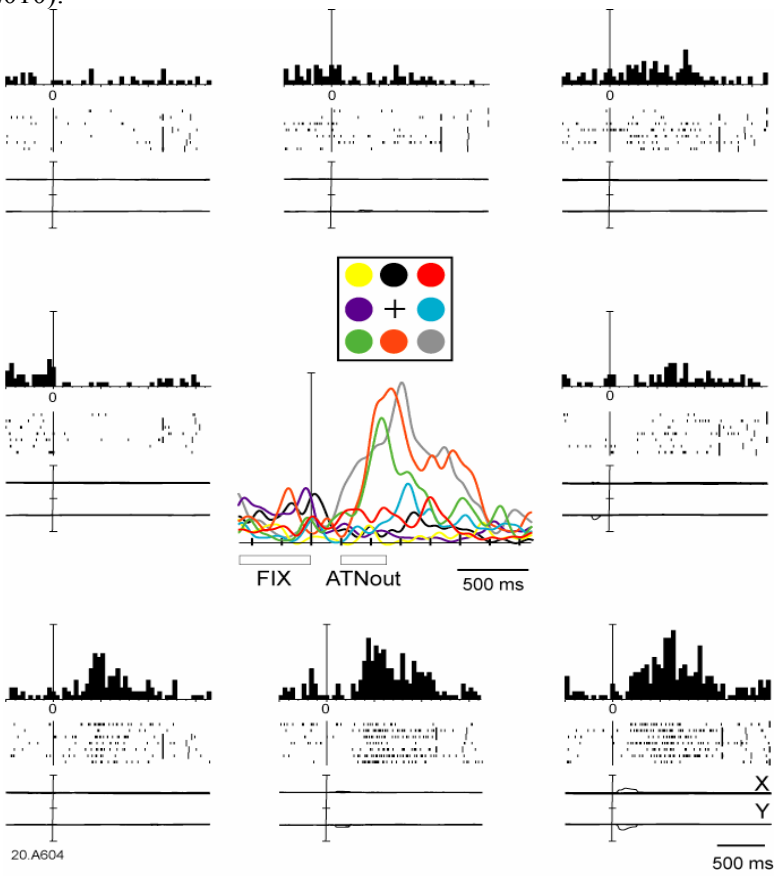


Figure 42: *Example of spatially-tuned modulations of neural activity during outward attention epoch. The neuron shows a strong discharge during outward attention epoch preferring covert shifts of attention towards the bottom part of the space. Each inset (positioned in the same relative position as the cue on the panel) contains the peri-event time histogram, raster plots and eye position signals. In the central part of the figure, the spike density functions (SDFs) of the activity for each of the 8 cue positions are superimposed and aligned on the cue onset. The mean duration of epochs FIX and outward attention is indicated below the SDFs. Neural activity and eye traces are aligned on the cue onset. Scalebar in peri-event time histograms, 70 spikes/s. Binwidth, 40 ms. Eyetraces: scalebar, 60°.*

This hypothesis that these attentional modulations may be helpful in guiding the hand during reach-to-grasp movements, particularly when the movements are directed to non-foveated targets is supported by the observation that often attentionl modulations occur in the same cells that show motor-related activity. This is the case for the cell reported in Fig. 43.
The example of Figure 43 shows that the effect of attention can modulate not only the ongoing activity but also the motor-related activity of single cells. The large majority of V6A cells are of this type. This is a cell whose activity was strongly modulated by the covert shift of attention towards the cue (outward attention epoch), but also by the action of button press, and by the bringing back of attention focus towards the fixation point (inward attention epoch). This last modulation was actually an inhibition. In addition, the cells shows also a transient visual response to the appearance of the peripheral cue. This example shows how V6A cells can be helpful in linking the fragment between vision and arm actions.

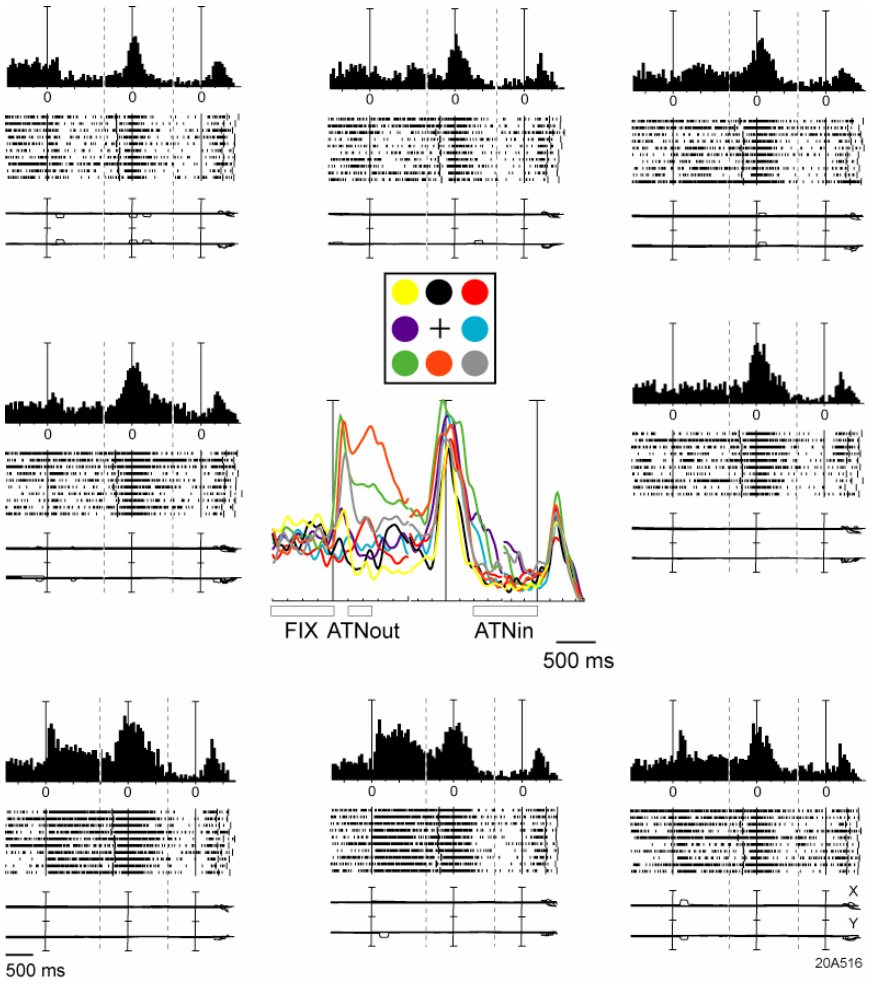


Figure 43: *Example of a cell modulated during outward and inward attention epochs and during arm actions (button release). This cell was excited during outward attention epoch when attention was covertly directed towards bottom locations, and inhibited during inward attention epoch for all attended locations. Neural activity and eye traces are aligned three times: from left to right: with the cue onset, with the button release and with the change in color of the fixation point. Peri-event time histograms: binwidth, 40 ms; scalebars, 180 spikes/s. Eyetraces: scalebar, 60°. All conventions are as in Fig. 42.*

UNIBO, together with partner WWU, suggested that area V6A is able to link perception to action in the 3D space through displacing the spotlight of attention. This area of the medial parieto-occipital cortex has anatomical connections (Gamberini *et al.,* 2009; Passarelli *et al.,* 2010 [C5]) that, together with these functional data, lead us suggest that V6A can provide to dorsal stream areas this information for producing the link across fragments by coordinating eye- and arm-actions in the 3D space.
Deliverable D5.2 reports more details on these data.

2) Encoding of 3D space through ocular movements

Link across single visual fragments can be obtained in many physiological situations. Commonly, in natural conditions, when we catch with vision a target of a potential reaching action, we move the eyes toward it and then the hand. Due to less inertia of the eyes, the eyes land on the target well before the hand starts to move. In area V6A of the medial parieto-occipital cortex, we have found neurons discharging in the epoch around the time of a saccade catching the visual target (perisaccadic epoch) and in the time of fixation, expecially the first 500 ms of fixation of a target in the dark. Interestingly, this kind of cells in V6A strongly prefer targets to be fixated in the peripersonal space, that is in the reachable space (Hadjidimitrakis *et al.,* 2010 [C3]).
In the third year of EYESHOTS, we demonstrated that area V6A of the medial parieto-occipital cortex of the macaque elaborates information related to directing the eyes to a visual target in depth. We explored ocular movements performed to targets located near the body up to positions located far away from the body, well beyond the reachable space. We found strong neural modulations related to changes of the vergence angle, a

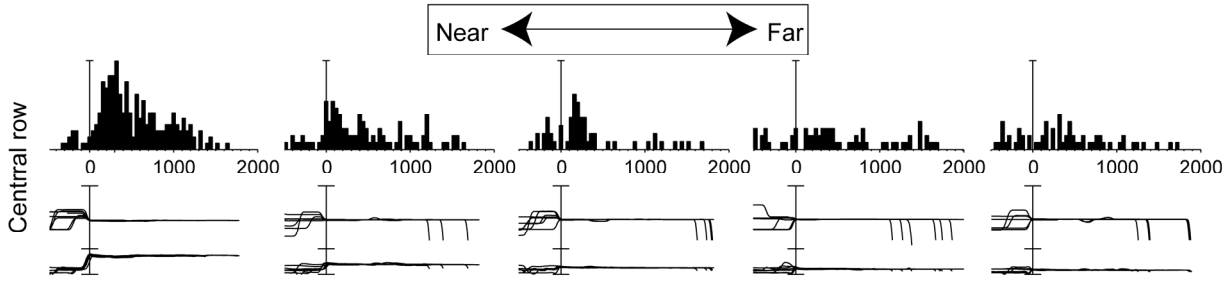


Figure 44: *Example of a neuron modulated in depth in the early fixation epoch.*
From top to bottom: neural responses and eye traces (version, top trace; vergence, bottom trace) to the five LEDs located on a midsagittal row, arranged from near (left) to far (right). The eye movement traces are aligned at the saccade onset. Scale bars for spike histograms and version and vergence traces were 80sp/s, 100 deg, and 20 deg, respectively.

result never reported so far for this cortical sector. Interestingly, neural discharges are stronger for oculomotor activity that brings fixation to targets located in the near space.

A neuron like the one reported in Fig. 44 is an example of such early-fixation signal, with a strong preference for near space. Together with a minority of V6A cells (15-35%) that showed a preference for the extrapersonal space and occasional units preferring intermediate distances, the large majority of V6A neurons (60-75%) preferred the peripersonal space, as also documented by the cell shown in Fig. 45.

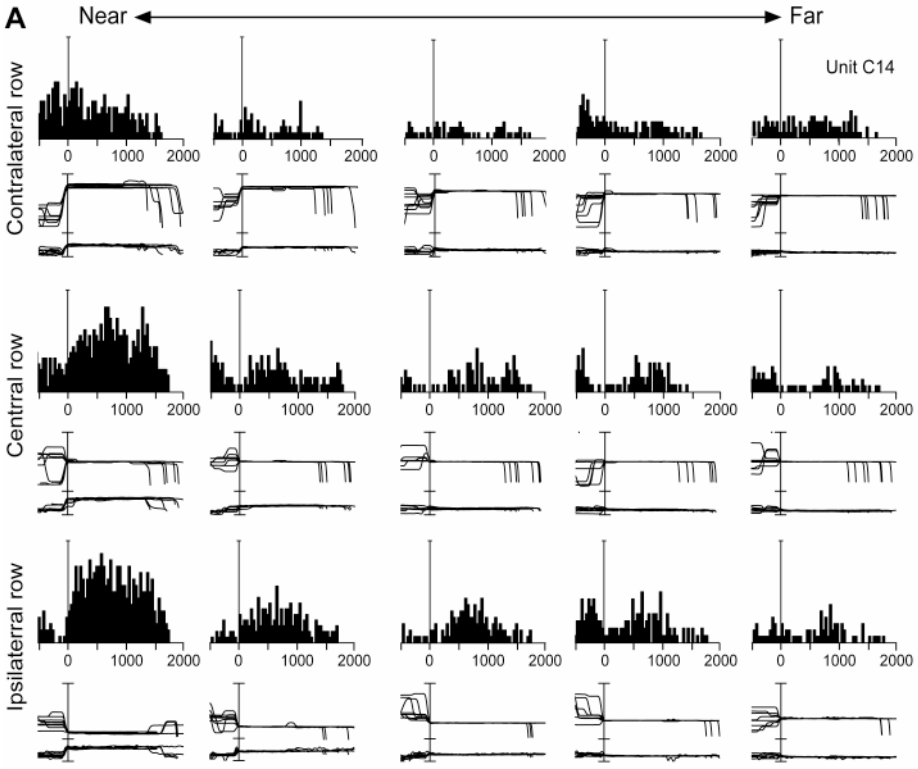


Figure 45: *Example of a neuron with fixation activity modulated by depth. Top/Middle/Bottom: neural responses and eye traces to the five LEDs of the contralateral/central/ipsilateral space arranged from near (left) to far (right), aligned at the end of the saccade. Scale bars for spike histograms and version (upper trace) and vergence (bottom trace) traces were 50 sp/s, 100 deg and 20 deg. Cell discharge reflects a strong tuning by vergence that is influenced by version, so that the cell is activated maximally for fixations on the nearest targets, especially in the ipsilateral space. All conventions are as in Fig. 1 and 3.*

The preference for near space at population level was also evident when we calculated the average normalized spike density function of each position in depth (Fig. 46). In each modulated cell, the activity of all LED positions were grouped together and averaged according to their distance from the animal (i.e. nearest, second near etc.), regardless of whether they were located in central or peripheral space. The cumulative activity of the cells modulated in perisaccadic epoch (left), aligned on saccade onset showed that the average responses were stronger for the nearest two targets (darkest lines). For the three farther LEDs neural activity was weaker and much less modulated. The same occurred for the cells modulated in the fixation period (right).
The perisaccadic activity found in V6A encodes the saccadic event, and this signal could be used to modulate the activity of arm reaching neurons. It is worthwhile to note that the information about eye position during the period around saccade is critical for the motor centers that controls the hand, because the retinal coordinates of the target change with the saccade. In this context, early fixation activity could constitute a fixation-for-reaching signal that brings the new retinal coordinates of the target to be grasped to the arm reaching neurons. The fact that V6A neurons, in both these intervals, clearly discharge more strongly in the reachable space, supports firmly this hypothesis.
We interpreted the neural encoding of V6A population as the neuronal correlate of a calibration between the eye and the arm systems and we proposed in the third year of the EYESHOTS project that the strong preference for reachable targets in early fixation period could reflect the shift of the attentional spotlight for the purpose of highlighting the location of the target of eye and hand movements in reaching an object (see Hadjidimitrakis *et al*., 2010 [C3]).

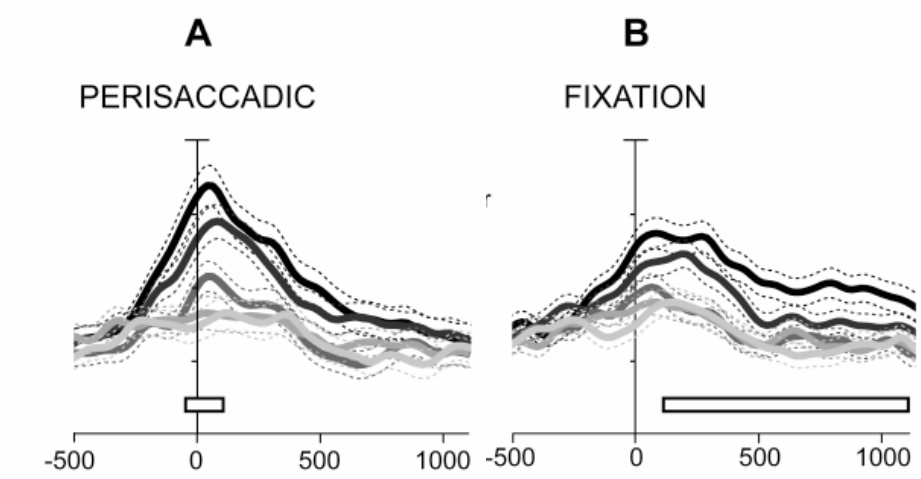


Figure 46: *Preference for near space at the population level. Population activity per each LED position (different tones of grey) of V6A cells modulated in (A) perisaccadic (N=132) and (B) fixation (N=193) considering all rows. Activity is expressed as averaged normalized SDF (thick lines) with variability bands (s.e.m., light lines) and is aligned at saccade onset in both (A) and (B). White rectangular boxes indicate the time intervals used at the permutation test (two nearest targets-the reachable ones- always different from the farthest ones, P<0.05, see text); vertical axis 10% of normalized activity per division and axis origin corresponds to 20% of normalized activity. By the time course of the population discharges it is evident that the 2 components (perisaccadic and fixation) are often coupled and that both of them preferentially encode the near space.*

UNIBO, together with UG, showed that many V6A neurons encode the spatial locations that the animal is going to gaze, or that it is actually fixating in 3D space, with a prevalence of cells prefering reachable locations. The abundance of strong tuning for near space could be a result of an adaptation process that balances the natural tendency for fixating far with the necessity to respond to behaviorally relevant stimuli appearing in the near space.
Deliverable D5.2 reports more details on these data.

Task 5.2: ***Link across fragments.***

Reaching movements in the 3D space.

The data collected in task 5.1 and summarized before suggest that V6A carries signals well suited to form a representation of the peripersonal/reachable space. This representation can be used to perform the sensori-to-motor transformations needed to perform successful reaching movements in depth.

The task designed for Task 5.2 and depicted in Fig. 47 is aimed at studying neural correlates of multisensory representation of 3D space obtained through active ocular and arm movements.
In the third year UNIBO performed analysis of electrophysiological data from 111 neurons from area V6A of the medial parieto-occipital cortex.
The task: 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 (MOV). The monkey has to push the target (HOLD), and to keep its hand on it until the fixation light switches off. The monkey releases the target and performs a backward movement (RET) toward the start button to be rewarded.

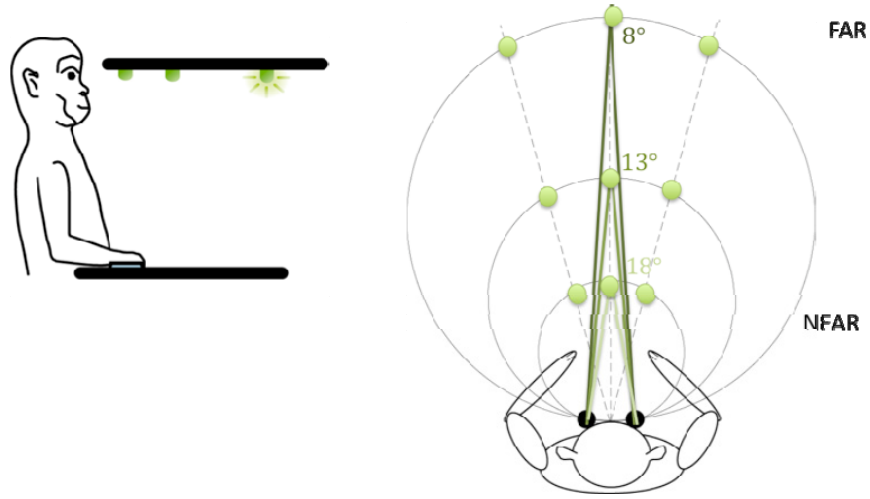


Figure 47: Lateral *view (left) and top view (right) of the reach-in-depth apparatus. Each green dot represents the target of a reaching movement. Vergence: 18-13-8°; version: -15-0-15°.*

Data Analysis: 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. MOV: from 200 ms before reaching movement onset to end of movement. HOLD: from end of reaching movement to 200 ms before onset of return reach. RET: from 200 ms before onset of return movement to return end.
Statistical tests: 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.
The study focused on the neural activity during arm movement/position-related epochs, that is during the execution of reaching movements toward and from targets located in a 3D space (MOV and RET) and during hand holding in these spatial locations (HOLD).
Since the neuronal encoding of reaching has been intensively studied by many laboratories in 2D space (frontal plane at fixed distance from the animal), we focused the present analysis on the reach-related modulations occurring in depth. We found that 57% of V6A neurons were modulated by depth in the epoch

MOV (63/111), 51% in the epoch HOLD (57/111) and 44% in the epoch RET (49/111). This means that several cells were influenced by depth in more than one arm-related epoch, as shown in the example of Fig.48. This cell showed a strong influence of depth on most of the time epochs we analysed. It was modulated by the saccade that brought the object on the fovea, during the fixation period (FIX), and during arm reaching related epochs (MOV, HOLD, RET). The cell shown in Fig. 48 had a coherent tuning of activity during fixation, reaching, hold, and return with increasing discharges for targets located in the far, left part of the space. This is particularly evident by comparing the discharges during MOV to targets far away (upper row) and to targets near the monkey (lower row) in each iso-version line (for example, the top left with the bottom left responses during MOV, and also during HOLD). It is worthwhile to note that during fixation, where no arm activity occurred, the neural modulations must be ascribed to the changes of vergence and version, particularly to the vergence, with a clear preference for low vergence angles. During reaching execution, the modulations in depth can be due to vergence influence on arm-motor signals or to reach-related signals influenced by arm-direction signals in the depth dimension.

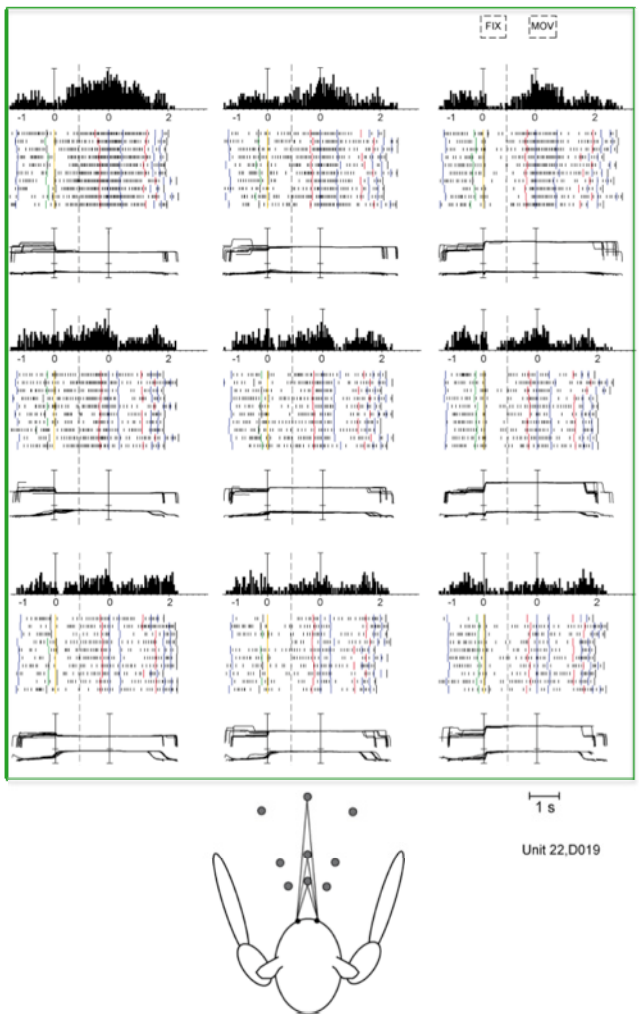


Figure 48: *V6A neuron modulated by depth of target locations in all arm reaching related epochs. The cell shows a spatial tuning, with a clear preference for reaches toward the far targets, especially the left ones. Alignement: saccade onset and reaching onset. All conventions are as in Fig. 1 and 4.*

In some cases, the tuning in depth of the neural discharges were confined to one arm reaching-related epoch only. The cell of Fig. 8, for instance, was modulated by the arm movement in depth only during forward movements (epoch MOV). The cell showed a spatial tuning. Neural activity in this case is enhanced when the monkey performed the reach toward the farthest targets. During fixation epoch, neither vergence nor version significantly modulated this cell.

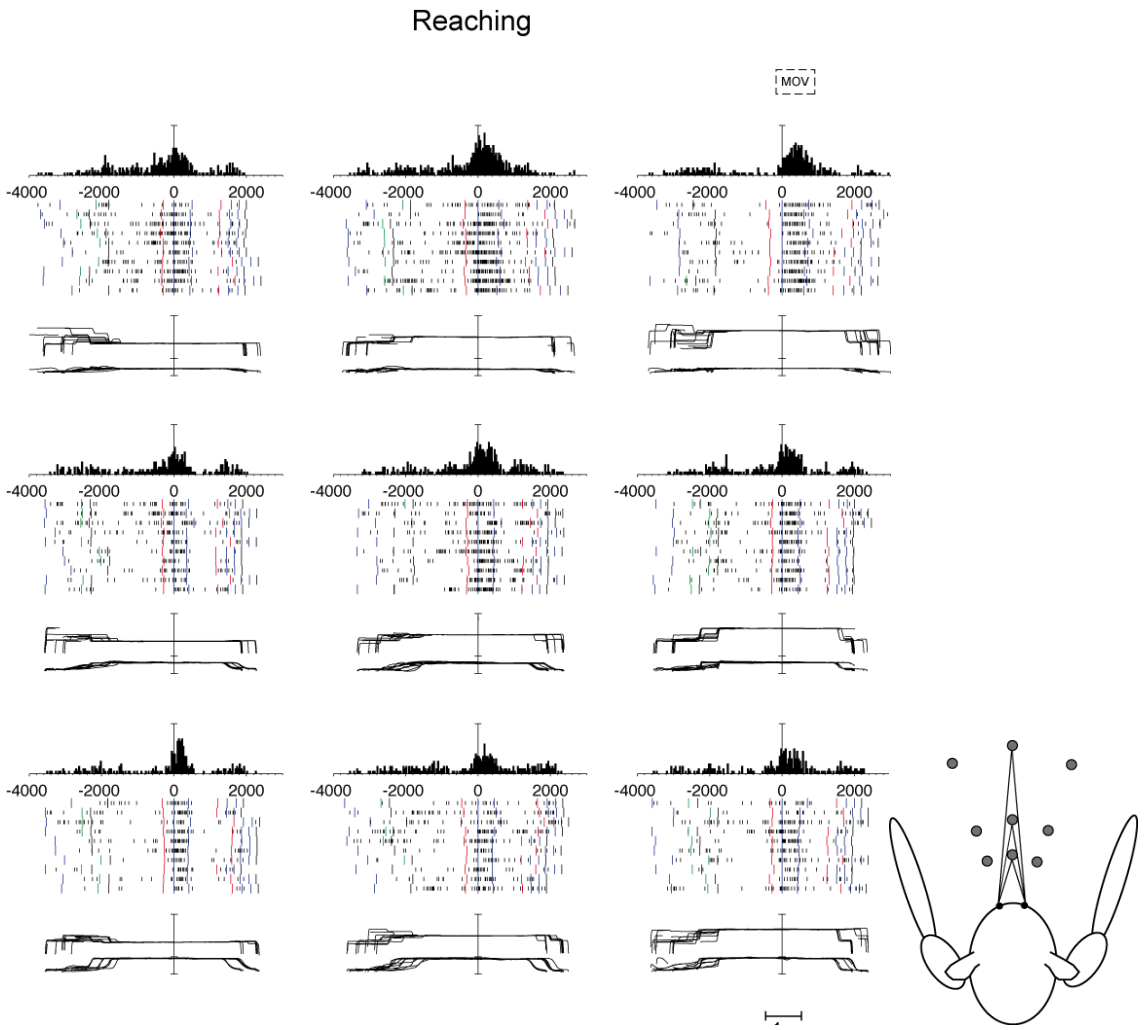


Figure 49: *Example of a V6A cell spatially tuned only for forward reaches in depth. The spatial modulation is specific for the MOV epoch, when the arm reaches its target. Activity is aligned on the onset of the reaching movement. All conventions are as in Figs. 42, 44, and 48.*

The ability of the entire neural population to discriminate the spatial position of the targets is demonstrated by the population spike density functions (SDFs) shown in the left part of Fig. 50. When the ranking of SDFs was based on the strength of neural activity (from best to worst activities of single cells), the curves were well apart one from another, with the worst discharge (blue curve) not different from the baseline activity, and the best discharge (violet curve) well above this level. This means that each cell strongly modulated its activity for different depths. In contrast, when the activity was ranked for all cells according to spatial location of the targets (Fig. 50, center) the 9 curves were superimposed, meaning that the cell population did not show a preference for a certain spatial position. This is also true when the targets were grouped according to their position in depth (Fig. 50, right): no preference for a given distance was observed. The ensemble of the plots shown in Fig. 50 demonstrates that, although the individual cells in V6A were tuned for reaching in depth (as shown in the examples reported in Fig. 49), the individual preferences compensated one another without a clear preference for a certain spatial location. In other words, the spatial fragments were sampled in the same way, with the same definition, across the entire reachable space.

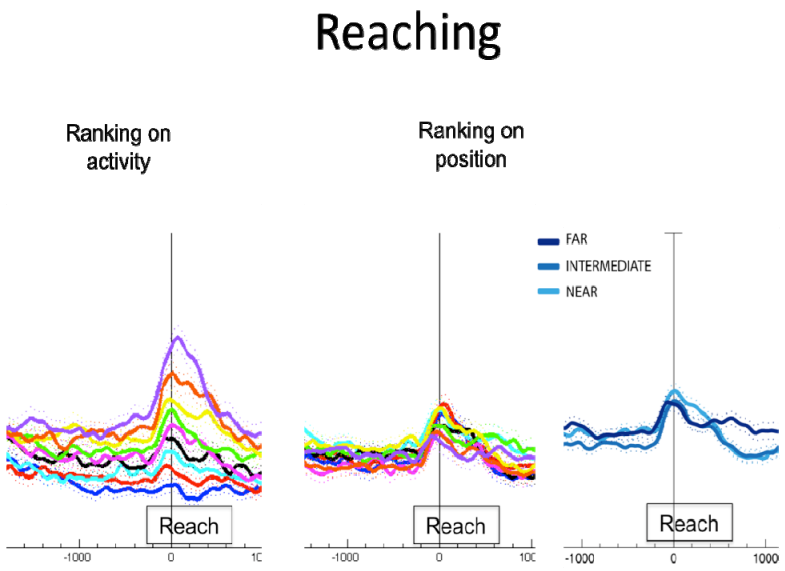


Figure 50: *Spatial tuning for reaches in depth at the population level.*
Population activity for each target position (different colors) of V6A cells modulated in the MOV epoch. Activity is expressed as averaged normalized SDF (thick colored lines) with variability bands (s.e.m., light lines) and is aligned at movement onset; vertical axis 100% of normalized activity. Rectangles labeled "Reach" indicate the mean duration of MOV epoch. More details are in the text.

Figure 51 shows the population discharges of cells tuned by return reach movements. As observed for forward reach movements, when ranking was based on the strength of neural activity, the SDFs were well apart one from another (Fig. 51, left), whereas when the activity was ranked according to the spatial location of target (Fig. 51, center, right) the curves were almost superimposed. This means, again, that the single cells were able to encode spatial locations, and the whole cell population encoded quite uniformly the 3D extrapersonal space.

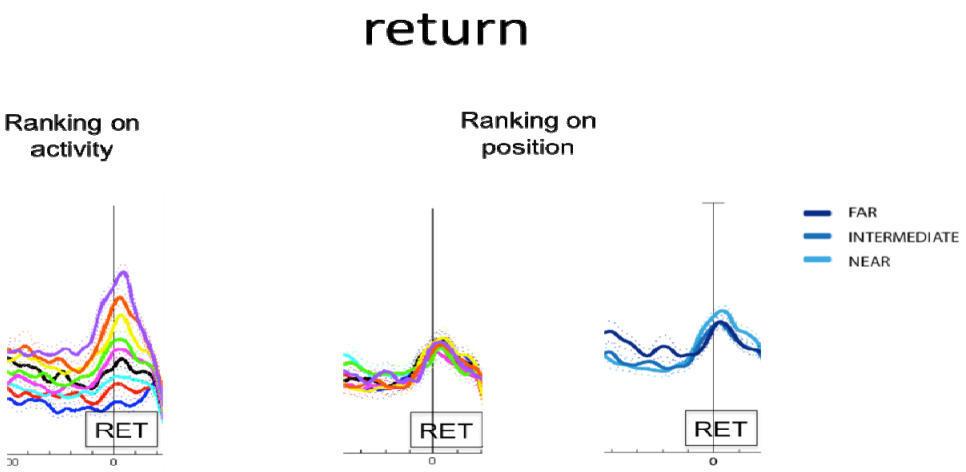


Figure 51: *Spatial tuning for return reaches in depth at the population level.*
Population activity for each target position (different colors) of V6A cells modulated in the RET epoch. Activity is expressed as averaged normalized SDF (thick colored lines) with variability bands (s.e.m., light lines) and is aligned at return movement onset; vertical axis 100% of normalized activity. White rectangles: mean duration of RET epoch. All conventions are as in Fig. 50.

Fifty-one per cent of V6A neurons (57/111) were modulated by depth in holding time. In this epoch, the monkey was keeping the hand immobile on the targets located in different spatial positions, at different depth in the peripersonal space.

The population as a whole (Fig. 52) showed the capacity to encode different spatial locations in holding time (ranking on activity; Fig. 52, left), but only slight preferences for the farthest depth or some positions in the 3D space (ranking on position; Fig. 52, center, right), similarly to what observed for reach-related epochs (Figs. 50 and 151). In other words, even for the static positions held on targets, the individual preferences of single neurons compensated one another without a clear preference for a certain spatial location. This means that the spatial fragments were sampled in the same way across the entire reachable space not only during the reaching epochs, but also during the holding time.

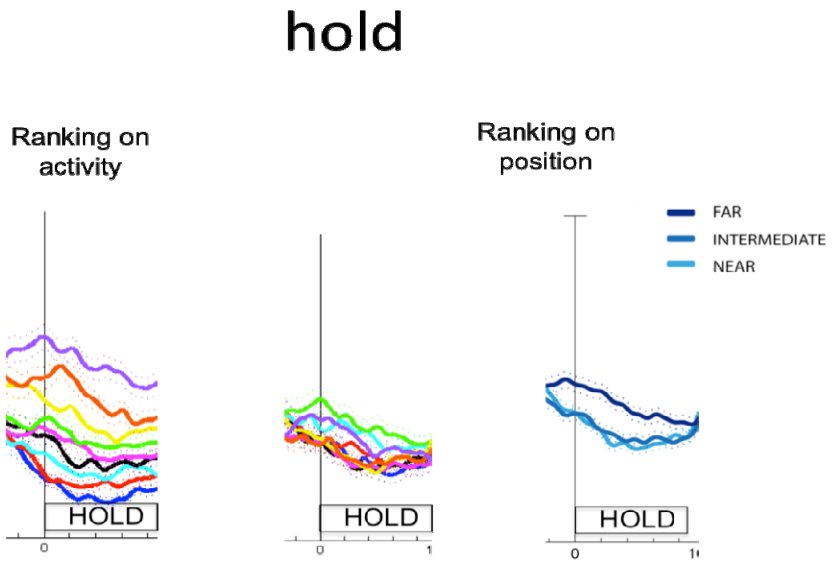


Figure 52: *Spatial tuning for HOLD in depth at the population level.*
Population activity for each target position (different colors) of V6A cells modulated in the HOLD epoch. Activity is aligned at holding time onset; vertical axis 100% of normalized activity. White rectangles: mean duration of HOLD epoch. All conventions are as in Fig. 50.

More analyses and details on these data are reported in deliverable D5.1(update).
A manuscript is in preparation with the results of this analysis (Breveglieri R, Hadjidimitrakis K, Bosco A, Sabatini S, Galletti C, Fattori P: Balanced sampling of visual fragments in the reachable space by parieto-occipital neurons, 2011).

Conclusions

All together, these functional data lead us to suggest that area V6A in the parieto-occipital cortex can provide to dorsal stream areas the necessary information for linking across visual fragments by coordinating eye- and arm-actions in the 3D space.
The present study suggests a novel role for the medial posterior parietal area V6A in constructing a 3D representation of the visuomotor world. It is known that V6A contains reach-to-grasp neurons (Galletti et al., 2003; Fattori et al., 2005; 2009; 2010) and cells that encode the two dimensional location of visual targets (Galletti et al., 1995), some of them in spatiotopic coordinates (Galletti *et al.*, 1993). Here we show that many V6A neurons also encode the spatial locations that the animal is reaching out, or has just reached out in the peripersonal space. Taken together, the data of the present study give strong support to the view that V6A plays a key part in the sensory-to-motor transformations that control reach-to-grasp arm movements and in elaborating sensory inputs and motor-like signals that could represent the internal body state for the purpose of sensorimotor integration.

Task 5.3: ***Motor description of fragment location***

This task is concerned with the role of motor parameters in fragment location. The perceived fragments in the peripersonal space are located via information based on the interaction of motor and visual parameters. The properties and condition of this coupling as well as its mechanism and parameters can be studied using the saccadic adaptation paradigm. In this paradigm a modification of the 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. During the first two periods of EYESHOTS we focused on the examination of the transfer of motor parameter changes to fragment location, and thus showed the contribution of motor preparation in localization. The results of these studies yielded indications that the initial eye position of a saccade had an effect on the process of saccadic adaptation. Up to now it has been assumed that saccadic adaptation was coded in a purely retinal reference frame. In that case the parameter of the initial eye position would not affect the calculation of motor parameters for saccade execution or the localization of targets. To examine the effect of the initial eye position on the process of saccadic adaptation and thus to specify the reference frame three studies were finished in the last period. The first one was a study with macaques which was accomplished in cooperation between the partners UNIBO and WWU. The second and third were studies with human subjects conducted at WWU. In our experimental setup the amplitude of saccades of monkeys and humans was modified by saccadic adaptation and afterwards the gain change of saccades of the same direction and amplitude but started in different positions was tested.

The *first study* with two participating monkeys was performed in the laboratory of UNIBO and a PhD student from WWU was participating in the stimulus presentation and data analysis. The experimental layout of the study is illustrated in Figure 53.

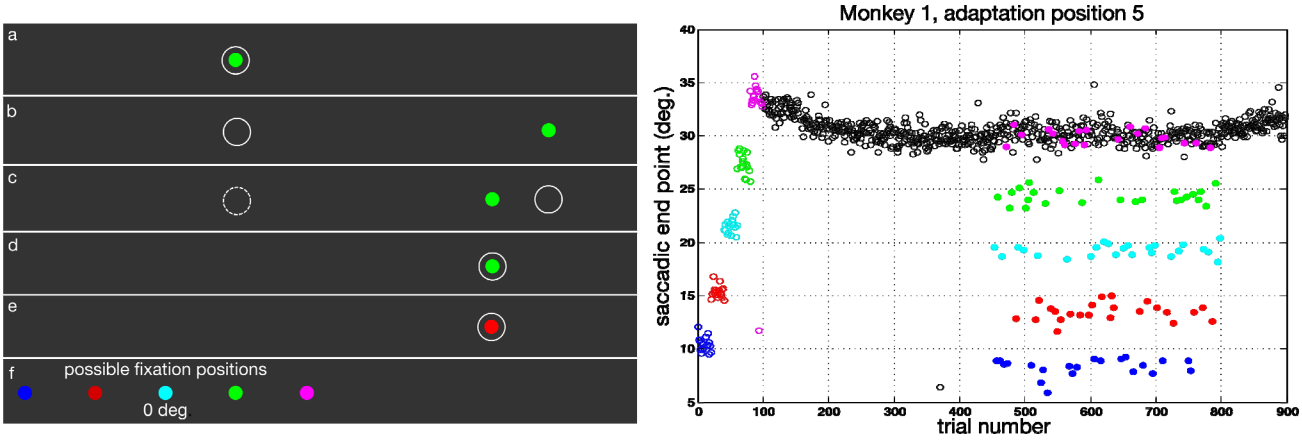


Figure 53: *Left panel: Sketches* ***(a)*** *–* ***(e)*** *show the experimental procedure for adaptation of reactive saccades.* ***a)*** *The green fixation point is presented and the monkey's gaze (circle) is directed towards it.* ***(b)*** *The fixation point is switched off and the green target appears.* ***(c)*** *At saccade onset the target is shifted.* ***(d)*** *The monkey makes a second saccade to land on the target.* ***(e)*** *After a randomized time the target becomes red and the monkey releases the button to get its reward. f) There were 5 possible starting positions of the saccade with a 6 deg spacing between adjacent points. Right Panel: The landing points of the saccades during an example session. The colors indicate the starting position of the saccades as shown under* ***(f)*** *and the black circles represent adaptation and de-adaptation trials (trial > 800) at the position + 12 deg.*

The two participating monkeys completed 5 sessions, each session with a different adaptation position (-12 deg., -6 deg., 0 deg., 6 deg., 12 deg.). The mean gain change measured at all 5 positions is presented in Figure 54 in percent of the applied target back step. The results show an unambiguous dependency of the amount of gain change in the test positions on the distance to the adaptation position. Hence, this data cannot be explained with a pure retinal reference frame but instead the eye position parameter needs to be incorporated. The observed gain modulation can be well described with a Gaussian transfer profile. This study is now being prepared for publication and will be presented at the European Converence on Eye Movements in July 2011.

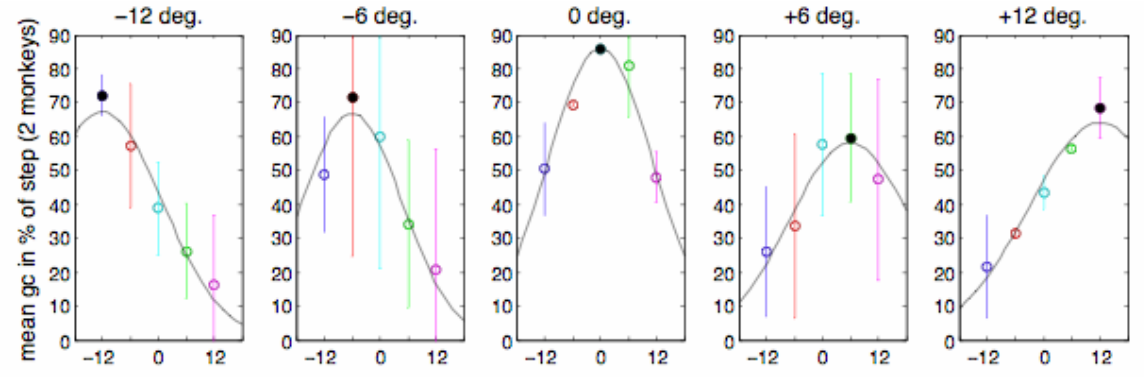


Figure 54: *Joined results of both participating monkeys. The circles show the mean gain change in each test position of every session. The error bars represent the standard error. The adaptation position of the displayed session is indicated by the filled circle. The data has been fitted with a Gaussian function (black line).*

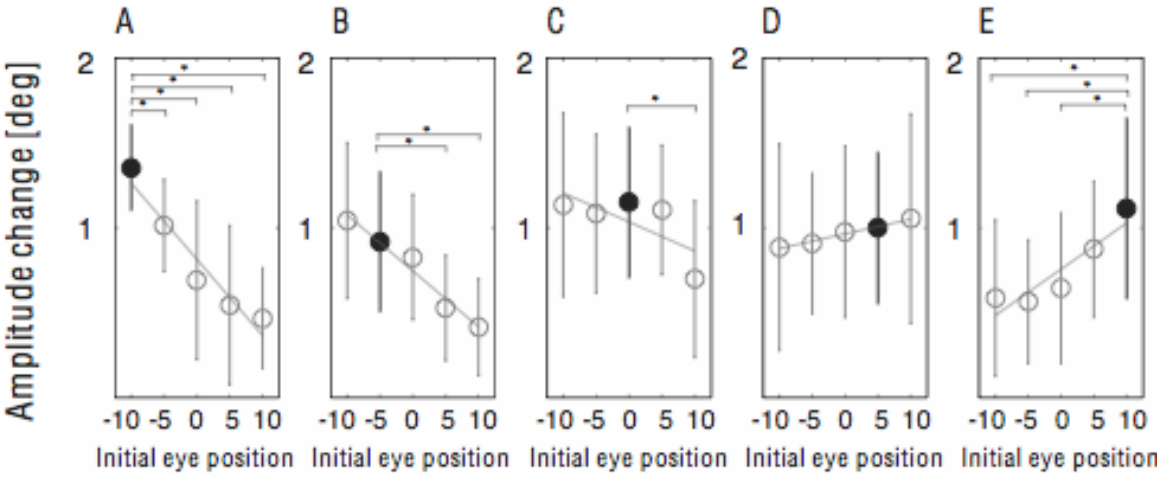


Figure 55: *Averaged amplitude changes for the horizontal arrangement of the different test positions. A-E) Each panel shows the amplitude changes for one adaptation session. The filled symbols show the adapted position in each session. A clear dependence of the amplitude change on the eye position is visible at the eccentric adaptation positions -10 deg, -5 deg, and 10 deg.*

The *second study* on the eye position effect in humans was performed with a very similar setup at WWU. Like in the study with monkeys there were 5 different initial eye positions of the saccades. The distance between two adjacent positions was 5 deg. from –10 deg. to +10 deg. In the first part of the study, the starting points were horizontally arranged. Figure 55 shows the group results of the 7 participating subjects. The mean amplitude change in a test position is affected by the location of the test position with respect to the adaptation position. These findings about a necessity of additional information beside the retinal reference frame are in line with the observations in the study with monkeys. But in contrast to the result of the monkey study, in humans a linear transfer profile well described the modulation of gain. The slopes of the linear fits are steep at the eccentric adaptation positions -10 deg, -5 deg, and 10 deg (Fig. 55A, B, and E), and shallow for positions 0 deg and 5 deg (Fig. 55C and D). The slopes at positions -10 deg, -5 deg and 10 deg are also significantly different from zero. This shows an increased influence of eye position for more eccentric adaptation positions in humans, whereas in monkeys the influence of the eye position remains constant over the tested excentricity. The study with human subjects had a second part in which the starting positions were vertically arranged instead of horizontally. The distance between 2 points remained at 5 deg and the saccades had the same vector like in the first part of the experiment. Like for the horizontal arrangement of initial position, the dependence of adaptation transfer on eye positions was strong in the most eccentric adaptation positions and shallow for the more central adaptation positions. These observations lead to the question of why adaptation at a central position in humans does not show an eye position dependent transfer whereas adaptation at an eccentric eye position does show a strong eye position effect. When considering eye position in saccadic adaptation, most approaches expressed eye position as context. One possibility to include eye position contexts into the mechanism of saccadic adaptation is an eye position

dependent modulation in a retinocentric reference frame (Fig. 56). Consider that neurons in many parts of the saccade circuitry encode space in a retinocentric reference frame and that the activity of these neuron is modulated by eye position gain fields. Then, for a given saccade vector, different neuronal subpopulations exist that fire more strongly for left or for right eye positions, respectively. Figure 4 depicts at the target representation stage in light gray a neuron pool preferring left eye positions, and in dark gray a neuron pool preferring right eye positions.

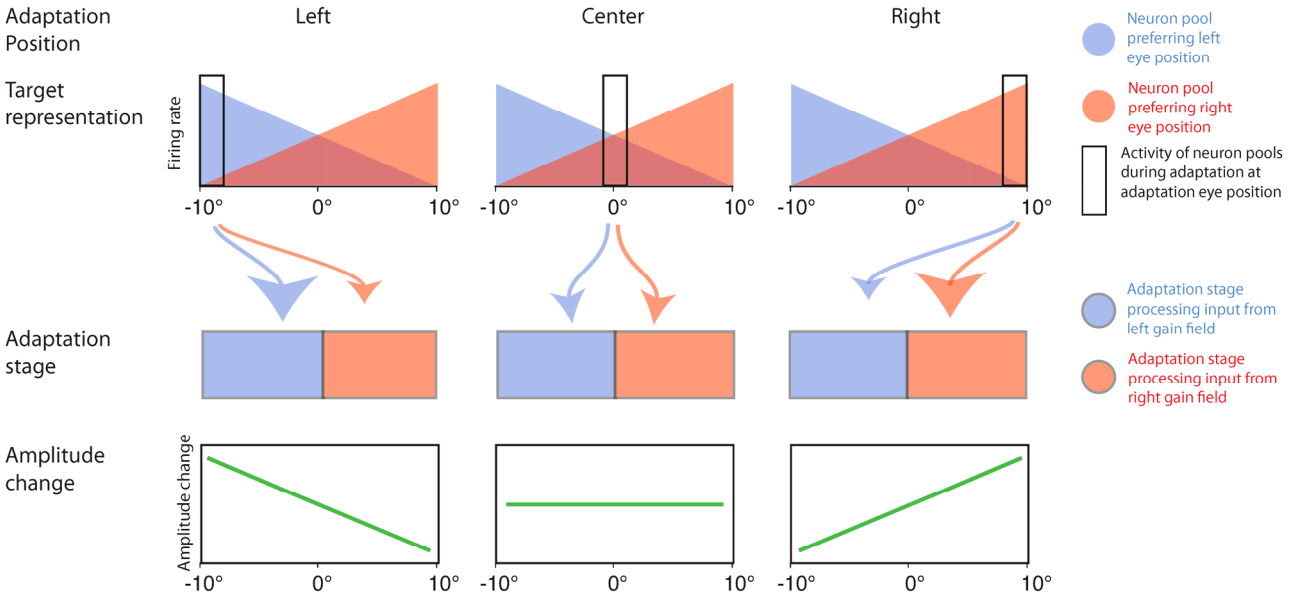


Figure 56: *Sketch of a possible mechanism for the eye position dependent modulation of saccadic adaptation.*

Depending on the initial eye position during adaptation, the two populations contribute differently to the generation of the saccade. If the activity of neurons with stronger saccade-related responses weighs more on the effects of adaptation, then mostly the left-preferring subpopulation contributes to the adaptation as shown by the size of the arrows to the adaptation stage in Figure 56. Saccades starting at right initial eye positions are driven mostly by the neuron pool shown in light gray, which is not adapted because it contributed little to the saccades originating from the adapted location. Therefore the amount of amplitude change will depend on initial eye position. However, when adapting at a central position, both subpopulations fire at intermediate rates, and both contribute to the saccade generation. Therefore, all neurons contribute to the adaptation and the amplitude change is seen at all eye positions. This study is under revision with the Journal of Neurophysiology.

The *third study* extended the investigation to the localisation of fragments. The transfer of gain change to the localisation of fragments was 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. Saccades of either type were adapted at one spatial location and gain change as well as mislocalisation of fragments was tested at that and three other spatial locations With this procedure we tested the reference frame of outward adaptation for reactive and scanning saccades and visual localization. For scanning saccades adaptation magnitude was drastically reduced at positions distant from the adapted eye position. Changes in visual localization showed a very similar modulation of eye position. These results suggest that scanning saccade adaptation is encoded in a non-retinotopic reference frame. Eye position effects for reactive saccade adaptation was smaller. No significant eye position specificity of mislocalization following reactive saccade adaptation was found. The findings reinforce earlier evidence that different reference frames are involved in reactive and scanning saccade adaptation and support the idea that oculomotor plasticity can occur at multiple sites in the brain. This study is now in press in the Journal of Neuroscience.

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. In the previous reporting periods we collected data

both in a single actor setting and on human-human interaction. In the current reporting period we conducted further data analysis and ran some control experiments. All results are reported in deliverable 5.4 and will be summarized below only briefly. After the planned experiments were finished, we added a new study that was concerned with human-robot interaction, specifically with the question how much the human accepts the robot as an interaction partner. Rather than asking the human subjects to rate the acceptability of the interaction we propose to use an unconscious behavioural measure, the *Social Simon Effect*.

Single actor setting experiments

The first two studies were single actor setting experiments conducted at the WWU in the first two reporting periods. For both experiments eye movements were measured with the Eyelink II eye tracker system (SR Research Ltd., Mississauga, Ontario, Canada). Study I investigated if gaze direction changes can be used to predict forthcoming pointing movements of another person. In sum, the results of Study I suggest that 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. Study II further explored the relation between gaze behaviour and arm movements and its influence on the allocation of attention. The results indicated that humans have a strong tendency to follow the gaze direction of another person, when this person simultaneously executes gaze and hand movements. The strong coupling between eye and hand movements could be used as a heuristic to distinguish relevant from irrelevant shifts in other's gaze direction. This study is currently under revision with Attention, Perception & Psychophysics

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. 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. A stereotypical gaze behaviour was observed: (1) at the start of each trial a fixation was directed towards the own object; (2) fixation was kept on a central location of the setup; (3) saccades were then regularly directed towards the partner's object in the terminal phase of the movement prior to the contact between objects; (4) the gaze followed the object until contact was made. In a second condition 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 case, the partner had to fully adapt his/her own movement to the trajectory of the first participant, and his/her predictability of the contact location was low. In order to compare the normal predictability and the low predictability conditions we focused on the timing of object directed saccades. Usually, the moment in time at which a saccade is directed towards a specific location can be used as an indicator of the relevance of that location in the execution of a specific task at that specific moment in time.

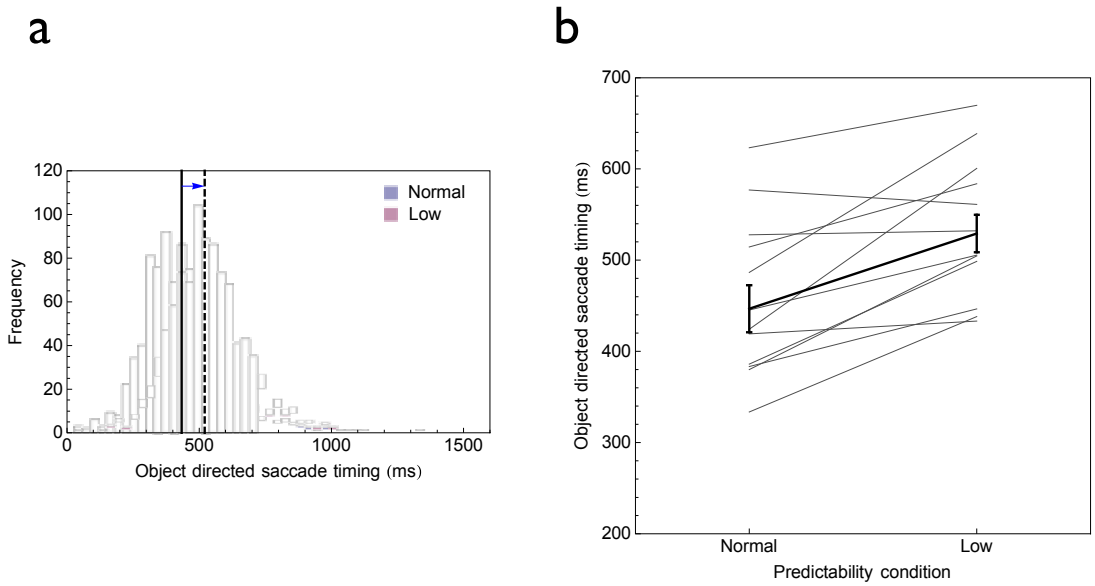


Figure 57: ***(a)*** *Distributions of the object directed saccade timings in the two conditions considered: 'normal' and 'low predictability'.* ***(b)*** *The difference in the object directed saccade timing between the*

normal and the low predictability conditions for all individual subjects (thin lines) and their mean (thick line).

Figure 57 shows distributions of the object directed saccade timings in the two conditions. The two distributions differ by a shift along the temporal axis. The timing of the object directed saccades was calculated with respect to the contact time between the two objects; therefore, it expresses how long before the contact a saccade was initiated towards the partner's object. The shift of the whole distribution of the low predictability condition thus indicates an earlier initiation of the saccades toward the partner's object. The difference in the object directed saccade timing between the normal and the low predictability conditions is represented in Figure 57(b) for all individual subjects and their mean.
A paired t-test showed that the object directed saccades started significantly earlier in the low predictability condition than in the normal predictability condition ($t(13)=2.505$, $P < 0.05$). The object directed saccades were initiated on average 529 ms (SD: 76 ms) before contact in the low predictability condition, and 446 ms (SD: 96 ms) before contact in the normal predictability condition. The difference is thus quite substantial considering that the whole trial, from the start of the object movement to the contact with the partner's object, lasted on average slightly more than 800 ms. On average, thus, participants performed the object directed saccade shortly after the start of object's movement (417 ms after the start in the normal predictability condition and 323 ms after the start in the low predictability condition). The necessity of gazing on the partner's object after this very short time lapse prevented the execution of any other gazing behavior toward other objects in the environment. When both participants were adapting their own hand movements to the trajectory of the other participant, they could start monitoring their partner behavior later in time than when only one participant had to take care of the whole adjustment by him/herself.
We further explored the relationship between the gaze behaviors of the two partners in the normal predictability condition by calculating the Pearson's product moment between the timings of the object directed saccades. The correlation coefficients ranged between 0.27 and 0.48. These correlations suggest that the partners were adjusting the timings of the object directed saccades with a certain degree of coordination, either both preceding of both delaying the moment in which they were performing the saccade toward the partner's object.
Our results thus suggest that the stereotypical gaze behaviour is 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 human-human interaction experiment met the milestone M9 at month 27 as planned. Deliverable D5.4 was submitted as planned. The study is currently prepared for publication.

<u>*Human-robot interaction*</u>

Additionally, a study on human-robot interaction were performed by members of the WWU in cooperation with the UJI lab. In this study we investigated under which conditions humans represent the actions of robots in a similar way to the actions of other humans, and thus accept the robot as a human-like interaction partner. Rather than assessing questionaires or ratings from the human subjects, which can be compromised by cognitive biases, we decided to develop an implicit behavioral test.
In human-human interactions, action co-representation plays a crucial role for successful joint action in shared workspace (Sebanz *et al.*, 2006). Action co-representation is typically investigated by using spatial compatibility tasks, like the *Simon Task*, in an interactive context (e.g. Sebanz *et al.*, 2003, 2005). In a *Simon Task* a person gives spatially defined manual responses to non-spatially stimulus attributes (e.g. the shape of a stimulus). Usually two stimuli are displayed on a monitor, either on the left or on the right side. The participant responds by pressing a key with either his/her left or right hand. Typically participants have to respond to the shape of a stimulus (the non-spatial stimulus attribute), but to ignore its location (the task-irrelevant attribute). Responses are usually faster when stimulus location and response location correspond, which is called the *Simon Effect* (Simon & Rudell, 1967). When a person only responds to one of the two stimuli (*Individual go/nogo task*), the Simon Effect disappears (e.g. Sebanz *et al*., 2003). However, the Simon Effect is re-established in the same go/nogo task when another person jointly responds to the complementary target stimulus (*Social Simon Task)*, which is called the *Social Simon Effect* (SSE). The SSE provides an index for action co-representation (e.g. Sebanz *et al.,* 2003; Tsai *et al.*, 2008).

Some studies seem to indicate that action co-representation is tuned to biological agents, hence facilitating human social interactions with conspecifics (e.g. Kilner, Paulignan, & Blakemore, 2003; Tsai & Brass, 2007; Tsai *et al*., 2008).
Recent studies on action observation may suggest that action co-representation may be explained by top-down attribution processes about the perceived animacy of the observed agent (Liepelt & Brass, 2010; Liepelt *et al.*, 2010).
Our study had two aims: First, we aimed to test if we can find evidence for action co-representation when interacting with a real robot that is perceived in a human-like way. Second, we aimed to test if the SSE can be used as a benchmark-tool for the perceived humanness of a robotic system.
For human-robot interactions in shared workspace, we hypothesized, that if action co-representation is purely biologically tuned, no SSE should be observed in a human-robot Social Simon Task, even when the robot is perceived as human-like. However, if action co-representation is sensitive to the perceived animacy of any given stimulus, even of a technical system, we should observe a SSE when interacting with a robot that is perceived as human-like.

We first investigated whether a SSE can be found when interacting with a real robot that is perceived as human-like. Therefore we developed an experimental setup in which a Social Simon Task was shared between a human and the UJI humanoid robot *Tombatossals* (see Figure 58a). 24 students of the UJI (12 male), aged 18 to 24 (mean age = 19.92 years, SD = 2.17) participated in Study IV. 13 of them were enrolled in technical studies, 11 of them were enrolled in humanities.
In the Social Simon Task, we used either a square or a diamond that was presented on the left or right side of the monitor. The human participant (seated to the left side of the monitor) responded to the square by pressing the left response key with the index finger of his/her right hand. The robot (seated to the right side of the monitor) responded to the diamond by pressing the right response key with the index finger of its left hand (see Figure 58b). The robot responded correctly on 98,4 % of trials, and performed errors in 1.6 % of trials. The experiment consisted of 512 trials in total. Prior to the experiment participants were told that the robot functions in a human-like manner. The robot's cameras were described as functioning similar to the eyes of a human, allowing the robot to actively detect small differences of the visual stimuli. The robot's behaviour was based on a biologically inspired neural network, thereby being able to decide when to respond. The robot was described as an active and intelligent agent.

(a) (b)

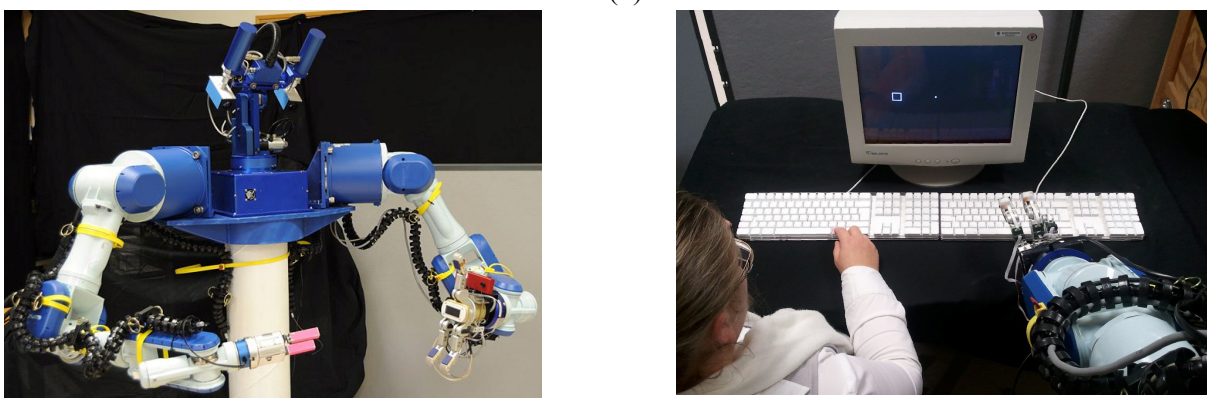


Figure 58: ***(a)*** *The figure shows the humanoid robot "Tombatossals" that was used for the human-robot interaction task.* ***(b)*** *The figure shows the experimental setup. The participant, seated on the left side of the monitor, responded to the square by pressing a button with the right hand. The robot, seated on the right side, responded to the diamond by pressing a button with its left hand.*

The analyses of reaction times showed a highly significant main effect of compatibility ($F(1,23) = 10.48$, $p = < .01$, $\eta^2 = .31$). Reaction times were significantly faster in compatible trials (340 ms) compared to incompatible trials (348 ms) (see Figure 59) clearly showing the Social Simon Effect in the order usually obtained for human-human interaction. This finding suggests that co-representation of non-biological agents can occur if an agent is perceived as human-like. The findings further indicate that the SSE may be used as a benchmark-tool for the perceived humanness of a robotic system. However, if this is true one should find no SSE when a robot functions purely deterministic and is perceived as non-human.

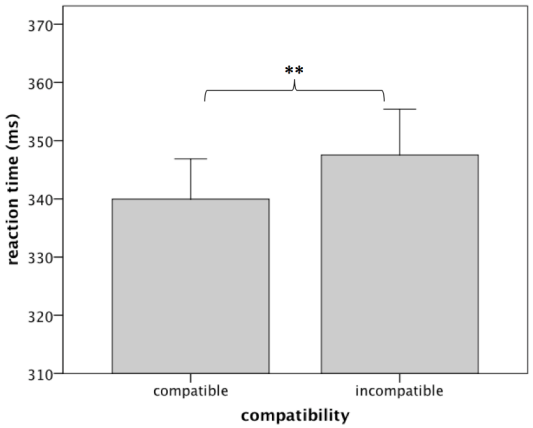


Figure 59: *Mean reaction times for compatible and incompatible trials in Study IV. Error bars represent the standard error of the mean. **: p < .01*

To test if it is actually the perceived humanness of the robotic system that induced the SSE, we conducted another experiment. We used the same Social Simon Task as before, but now we introduced the robot as functioning in a non-human, and purely deterministic manner. The robot was described as a mechatronic device, which movements were controlled by electrical motors. Participants were told that the robot's behaviour was fully determined by the commands of a computer program, which sent an actuation signal to the motors of the robot hand every time it should respond. Like this, the robot was described as a purely deterministic agent, which passively executes external commands.
A new group of 24 students of the University of Castellon (12 male), aged 18 to 38 (mean age = 20.37 years, SD = 4.31) participated in the study. 15 of them were enrolled in technical studies, 9 of them studied humanities. The reaction time analyses revealed no main effect of compatibility ($F(1,23) = 1.48$, $p>0.05$, $\eta^2=0.06$). Reaction times in compatible trials (333 ms) did not differ statistically from the reaction times in incompatible trials (336 ms) (see Figure 60).

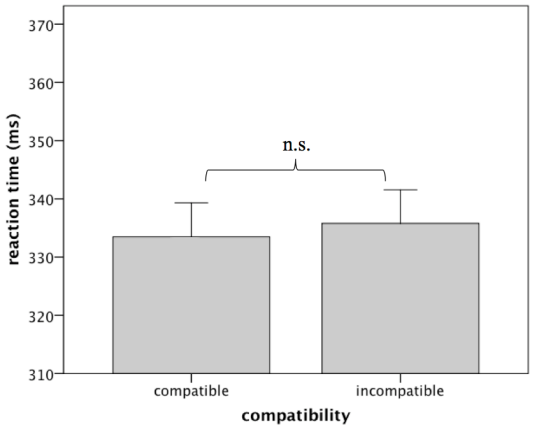


Figure 60: *Mean reaction times for compatible and incompatible trials in Study V. Error bars represent the standard error of the mean. n.s.: p > .05*

Since no SSE occured in this experimment, we conclude that action co-representation as measured by the SSE does not seem to occur when interacting with a purely deterministic and non-human-like robot.
Taken together, our results indicate that action co-representation is not exclusively tuned to biological agents. Instead, higher order cognitive processes, like the perceived humanness of a robot, seem to influence the amount of action co-representation in a top-down manner. Co-representation of robotic (non-biological) actions can occur if a robot is perceived as functioning in a human-like way.
The presently developed experimental setup measuring the SSE in human-robot interactions seems to provide a good measurement to indicate the perceived humanness of a robotic system. The study is currently prepared for publication (Stenzel *et al.,* 2011).

Deviations from the project workprogramme

None.

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
D4.2b	Autonomous generation of object awareness	WP4	UJI	D	PU	27	Yes	04-Jun-10 / 31-May-10	None
D5.4	Report on cooperative behaviour in shared workspace	WP5	WWU	R	PU(*)	27	Yes	07-Jun-10 / 31-May-10	None
D1.2 (update)	Non-visual depth cues and their influence on perception	WP1	UG	R	PU	30	Yes	07-Sep-10 / 31-Aug-10	None
D1.3	Control of voluntary transfer of fixations to new depth planes	WP1	UG	O	PU(*)	30	Yes	06-Oct-10 / 31-Aug-10	None
D1.4	Bioinspired Stereovision Robot System	WP1	UG	P	PP	30	Yes	15-Sep-10 / 31-Aug-10	A revised version of the supporting documentation (April 2011) is available, which includes details on the final prototyping activity.
D2.1 (update)	Convolutional network for vergence control	WP2	K.U.Leuven	R	PU	30	Yes	15-Sep-10 / 31-Aug-10	None
D3.2	Object-based top-down selection	WP3	WWU	O	PU(*)	30	Yes	03-Sep-10 / 31-Aug-10	None
D4.3a	How to build a global awareness of the peripersonal space	WP4	UJI	R	PU(*)	30	Yes	08-Sep-10 / 31-Aug-10	None

D5.2	Report on monkey eye movements and arm movements in the link across fragments	WP5	UNIBO	R	PU(*)	30	Yes	07-Sep-10 / 31-Aug-10	None
D5.3b	Report on the respective influence of motor and visual parameters on fragment location obtained from saccade adaptation data on monkeys	WP5	UNIBO	R	PU(*)	30	Yes	15-Sep-10 / 31-Aug-10	None
D2.2b	Algorithm for 3D scene description through interactive visual stereopsis adaptation using the mechatronic system	WP2	K.U.Leuven	D,R	PU(*)	36	Yes	23-Feb-11 / 28-Feb-11	None
D3.3b	Final, fully tested version of the Working Memory Model	WP3	WWU	R	PU(*)	36	Yes	25-Mar-11 / 28-Feb-11	None
D4.3b	An embodied agent which learns to situate itself in the environment through active exploration	WP4	UJI	D	PU	36	Yes	21-Mar-11 / 28-Feb-11	None
D4.3c	Final robot head-eye/arm set-up featuring the robot eye system developed within WP1	WP4	UJI	D	PU	36	Yes	21-Mar-11 / 28-Feb-11	None
D5.1 (update)	Report on neural discharges in the medial parieto-occipital cortex	WP5	UNIBO	R	PU(*)	36	Yes	24-Feb-11 / 28-Feb-11	None
D5.4 (update)	Report on cooperative behaviour in shared workspace	WP5	WWU	R	PU(*)	36	Yes	24-Feb-11 / 28-Feb-11	None
D8.2	A collection of one page student reports about cooperation work	WP8	UG	R	CO	36	Yes	21-Mar-11 / 28-Feb-11	None

(*) 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**
M9	Experimental data on human-human interaction obtained	WP5	WWU	27	Yes/No	May 2010	We obtained and analyzed the data from the human-human interaction experiemnts. The results confirm our hypothesis that that gaze tracking can be used to predict cooperative behavior
M10	Eye position gain fields	WP1	UG	30	Yes/No	August 2010	We have demonstrated that the responses of the population of disparity detectors can be profitably adapted according to the current position of the eyes. Gaze information can be used as a prior to reallocate the resources, e.g. by redistributing the cells of the population, or by modulating their responses. The milestone has been reached as planned.
M11	Algorithm for robust head-centric 3D description of visual fragments	WP2	K.U.Leuven	30	Yes/No	August 2010	We have developed both computer vision and biologically plausible approaches that can directly transform the disparity detectors' population response into a head-centric representation for accuracte 3D description of the fragment. Robustness to the limited accuracy of the motor system has been demonstrated as well.

M12	Arm reaching gain fields	WP4	UJI	30	Yes/No	August 2010	The robot system that implements the neural network model framework inspired by V6A data and concepts is able to build a perceptual awareness of objects in its peripersonal space and to perform concurrent or decoupled reach and gaze movements toward them.
M13	Anthropomorphic eye system	WP1	UG	30	Yes/No	August 2010	First prototype released and experimentally tested. Ocular working range of 90^o. Ocular angular speed in excess of 90^o/*sec* (as of servo motor data sheet specifications) Eyeball diameter 28 mm. Embedded 5Mpixels USB camera. Experimentally tested closed loop bandwidth 5Hz (unfortunately determined by the commercial servo amplifiers adopted). Accurated design to respect Listing's Law (*double inverted torus* design for custom dry bearing used to support the Eyeball).
M14	Interactive stereopsis system	WP2	K.U.Leuven	36	Yes/No	February 2011	We have demonstrated the effectiveness of the developed robust algorithms for obtaining head-centric 3D information by applying them on real-world images obtained with the iCub stereo head.
M15	Construction of a global awareness of the peripersonal space	WP4	UJI	36	Yes/No	February 2011	As planned, the system is able to build a multisensor egocentric 3D representation of a simplified natural environment involving a set of everyday life objects, as the result of its multimodal interaction in its peripersonal space. On request it is able to change the gaze or reach towards a particular object while keeping a continuous visual exploration of the scene.

9 References

Alexander, G.E., DeLong, M.R. and Strick, P.L., Parallel organization of functionally segregated circuits linking the basal ganglia and cortex. *Annual Review of Neuroscience,* **9:**357-381, 1986.

Antonelli, M., Chinellato, E., del Pobil, A.P. Implicit mapping of the peripersonal space of a humanoid robot. *IEEE Symposium Series on Computational Intelligence - SSCI 2011*, Paris, April 2011.

Beuth, F., Wiltschut, J. and Hamker, F.H., Attentive Stereoscopic Object Recognition, in: *Machine Learning reports of AG Computational Intelligence, University of Leipzig,* **04/2010**:41-48, 2010.

Biamino, D., Cannata, G., Maggiali, M. and A. Piazza. Mac-eye: a tendon driven fully embedded robot eye. *In 5th IEEE-RAS International Conference on Humanoid Robots*, pages 62–67, 2005.

Bosco A., Breveglieri R., Chinellato E., Galletti C. and Fattori P. Reaching activity in the medial posterior parietal cortex of monkeys is modulated by visual feedback *J. Neurosci*. 30:14773-85, 2010

Brown, J.W., Bullock, D. and Grossberg, S., How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks* **17**:471-510, 2004.

Canessa A., Sabatini S.P., Solari F. Visuo-motor constraints in binocular eye coordination: optimization theories revisited. Submitted.

Cannata, G. and Trabucco, A. Coordinated Tendon Control of a Bioinspired Robot Eye", *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011)*, submitted.

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

Chessa M., Solari F., Sabatini S.P. A Virtual Reality Simulator for Active Stereo Vision System. *International Conference on Computer Vision Theory and Applications 2009*,VISAPP'09, Lisbon 5-8 February 2009a.

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, 2009b

Chessa, M., Solari, F. and Sabatini, S.P. Virtual Reality to Simulate Visual Tasks for Robotic Systems, *Virtual Reality*, Jae-Jin Kim (Ed.), ISBN: 978-953-307-518-1, InTech, 2011

Chinellato, E., Grzyb, B.J., Marzocchi, N., Bosco, A., Fattori, P. & del Pobil, A.P. The Dorso-medial visual stream: From neural activation to sensorimotor interaction. *Neurocomputing*, **74**(8):1203-1212, 2011a.

Chinellato E., Antonelli, M., Grzyb, B.J. del Pobil, A.P. Implicit sensorimotor mapping of the peripersonal space by gazing and reaching. *IEEE Trans. on Autonomous Mental Development*, **3**(1):43-53, 2011b.

Chinellato E., Felip J., Grzyb B.J., del Pobil A.P. Hierarchical object recognition inspired by primate brain mechanisms. *IEEE Symposium Series on Computational Intelligence - SSCI 2011*, Paris, April 2011c.

Chown, E. Making predictions in uncertain world: Environmental structure and cognitive maps. *Adaptive Behavior* 1999, **7**(1):17-33

Chumerin, N., Gibaldi, A., Sabatini, S.P., Van Hulle, M.M. Learning eye vergence control from a distributed disparity representation. *International Journal of Neural Systems*, **20**(4):267-278, 2010

Clark, M. and Stark, L. Time optimal behavior of human saccadic eye movement. *IEEE Transaction on Automatic Control*, **20**(3):345-348, 1975

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

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.

Enderle, J. D., Wolfe, J. W. and Yates, J. T. The linear homeomorphic saccadic eye movement model- A modification. *IEEE Trans. Biomed. Eng*., **31**(11):717-720, 1984.

Fattori, P., Raos V., Breveglieri R. , Bosco A. , Marzocchi N. , Galletti C. The dorsomedial pathway is not just for reaching: Grasping neurons in the medial parieto-occipital cortex of the macaque monkey *J Neurosci*, **30**: 342-349, 2010

Fattori, P, Breveglieri R, Marzocchi N, Filippini D, Bosco A, Galletti C. Hand orientation during reach-to-grasp movements modulates neuronal activity in the medial posterior parietal area V6A. *J Neurosci.* **29**:1928-36, 2009

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.

Fleet, D.J., Wagner, H. and Heeger. D.J. Modelling binocular neurons in the primary visual cortex. In *Computational and Biological Mechanisms of Visual Coding*, M. Jenkin and L. Harris (eds.), Cambridge University Press, pp. 103-130, 1996.

Földiak, P. Learning invariance from transformation sequences. *Neural Computation*, **3**:194-200, 1991.

Fuke, S.; Ogino, M. & 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., Fattori, P. Parietal neurons encoding spatial locations in craniotopic coordinates. *Exp Brain Res* **96**:221-229, 1993.

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 *PLoS ONE* **5**(11): e15078. doi:10.1371/journal.pone.0015078, 2010.

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., Breveglieri, R., Bosco, A., Fattori, P. and Galletti, C. Functional profiles within the medial posterior parietal area V6A. *Program No. 75.6. 2010 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2010.*

Gamberini, M., Galletti, C., Bosco, A., Breveglieri, R. and Fattori, P. Is the medial posterior-parietal area V6A a single functional area? *J Neurosci*. 31: 5145-5157, 2011

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.

Gibaldi A., Canessa A., Chessa M., Sabatini S.P., Solari F. Read-out rules for short-latency disparity- vergence responses from populations of binocular energy units: the effect of vertical disparities Submitted to 33th European Conference on Visual Perception (ECVP'10), Lausanne, 22-26 August, 2010.

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.

Haber, S.N., The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy,* **26:**317-330, 2003.

Hadjidimitrakis, K., Breveglieri, R., Bosco, A., Placenti, G., Bertozzi, F., Fattori, P. and Galletti C. Fixation in depth reveals a preference for reachable space in the macaque medial parieto-occipital cortex. *Program No. 75.7. 2010 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2010.*

Hadjidimitrakis, K., Breveglieri, R., Placenti, G., Bosco, A., Sabatini, S.P., Galletti, C. and Fattori, P. Fix Your Eyes Where You Can Reach: Neurons in the Macaque Medial Parietal Cortex Prefer Gaze Positions in Peripersonal Space, 2011. Submitted

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

Hansen, M. and Sommer, G. Active depth estimation with gaze and vergence control using gabor filters. In *Proceedings of the 13th International Conference on Pattern Recognition,* , **1**:287–291, 1996.

Havermann, K., Zimmermann, E., Fattori, P. and Lappe, M.Eye position effects in the adaptation of reactive saccades *532.2 2010 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2010*

Hibbard, P.B. A statistical model of binocular disparity. *Visual Cognition*,**15**(2):149–165, 2007

Hollerman, J.R. and Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience,* **1**(4):304-309, 1998.

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.

Horvitz, J.C., Stimulus-response and response-outcome learning mechanisms in the striatum. *Behavioural Brain Research,* **199**(1):129-140, 2009.

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

Hulse, M.; McBride, S.; Law, J. & Lee, M. Integration of Active Vision and Reaching From a Developmental Robotics Perspective. *IEEE Transactions on Autonomous Mental Development*, **2**:355-367, 2010.

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.

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

Kilner, JM, Paulignan, Y, Blakemore S-J. An interference effect of observed biological movement on action. *Current Biology* **13**(6): 522-525, 2003.

Krishnan, V.V. and Stark, L.A. A heuristic model for the human vergence eye movement system. *IEEE Trans. Biomed. Eng.*, **24**:44-49, 1977.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE,* **86**(11):2278-2324, 1998.

Lenz, A., Anderson, S.R., Pipe, A.G., Melhuish, C., Dean P. and Porrill J., Cerebellar-Inspired Adaptive Control of a Robot Eye Actuated by Pneumatic Artificial Muscles. *IEEE Trans. on Systemes Man and Cybernetics Part B Cybernetics*, **39**(6): 1420-1433, 2009.

Liepelt, R., and Brass, M. Top-Down modulation of motor priming by belief about animacy. *Exp Psychol*, **57**:221-227, 2010.

Liepelt, R., Prinz, W., and Brass, M. When do we simulate non-human agents? Dissociating communicative and non-communicative actions. *Cognition,* **115**:426-434, 2010.

Lindeberg, T. and Florack, L. Foveal scale-space and the linear increase of receptive eld size as a function of eccentricity. *Technical report ISRN KTH NA/P-94/27-SE. 1994*

Liu, Y., Bovik, A.C. and Cormack, L.K. Disparity statistics in natural scenes. *J. of Vision*, **8**(11):1–14, 2008

Marfil, R., Urdiales, C., Rodrıguez, J. A. and Sandoval, F. Automatic Vergence Control Based on Hierarchical Segmentation of Stereo Pairs. *International Journal of Imaging Systems and Technology,* **13**(4), 224–233, 2003

Marjanovic, M.; Scassellati, B. and Williamson, M. Self-taught visually-guided pointing for a humanoid robot. *International Conference on Simulation of Adaptive Behavior* (SAB) 1996.

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.

Mehmood, A., Camescasse, B., Ouezdou, F.B. and Cheng G. Simulation and Design of 3-DOF Eye Mechanism Using Listing's Law. 8TH IEEE-RAS *International conference on Humanoid Robots.* 444-449, 2008.

Nori, F.; Natale, L.; Sandini, G. and Metta, G. Autonomous learning of 3D reaching in a humanoid robot. *IEEE International Conference on Intelligent Robots and Systems*, 2007.

Nosengo N.. Robotics: The bot that plays ball. *Nature,* **460**:1076-1078, 2009

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 Computation,* **18**:283-328, 2006.

Passarelli, L., Gamberini, M., Bakola, S., Burman, K. J., Fattori, P., Rosa, M.G. and Galletti, C. Cortico-cortical connections of the parietal area V6Av in the macaque monkey. *Program No. 373.21. 2010 Neuroscience Meeting Planner. San Diego, CA: Society for Neuroscience, 2010.*

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

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

Pouget, A., and Sejnowski, T.J. Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*. **9**(2):222-237, 1997

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.

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

Redgrave, P., Rodriguez, M., Smith, Y., Rodriguez-Oroz, M.C., Lehericy, S., Bergman, H., Agid, Y., DeLong, M.R. and Obeso, J.A., Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. *Nature Reviews Neuroscience,* **11**(11):760-772, 2010.

Reynolds, J.N., Hyland, B.I. and Wickens, J.R., A cellular mechanism of reward-related learning. *Nature,* **413:**67-70, 2001.

Rolls, E.T. and Stringer, S.M. Invariant object recognition in the visual system with error correction and temporal difference learning. *Network*, **12**(2):111-129, 2001.

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*, **114**(6):681–699, 2010.

Scharstein, D., Szeliski, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Journal of Computer Vision* **47**(1/2/3): 7–42, 2002

Schenck, W.; Hoffmann, H. & Möller, R. Learning Internal Models for Eye-Hand Coordination in Reaching and Grasping. *Proc. EuroCogSci*, 289-294, 2003.

Schnier, F., Zimmermann, E. and Lappe M. Adaptation and mislocalization fields for saccadic outward adaptation in humans. *Journal of Eye Movement Research*, **3**(3):1–18, 2010.

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

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

Schroll, H. Vitay, J., Hamker, F.H. Working memory and response selection: A computational account of interactions among cortico-basal ganglio-thalamic loops, Submitted to *Journal of Congnitive Neuroscience*

Schultz, W., Dayan, P. and Montague, P.R., A neural substrate of prediction and reward. *Science,* **275**:1593-1599, 1997.

Schultz, W., Getting formal with dopamine and reward. *Neuron,* **36**:241-263, 2002.

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

Sebanz N., Knoblich, G., and Prinz, W. Representing others‘ actions: just like one‘s own? *Cognition*, **88**:B11-B21, 2003.

Sebanz, N., Bekkering, H., and Knoblich, G. Joint action: Bodies and minds moving together. *Trends Cogn Sci*, **10**:70-76, 2006.

Sebanz, N., Knoblich, G., and Prinz, W. How to share a task: Corepresenting Stimulus-Response mappings. *J Exp Psychol Human*, **31**:1234-1246, 2005.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. and Poggio, T. Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.*, **29**:411-426, 2007.

Shen, W., Flajolet, M., Greengard, P., and Surmeier, J., Dichotomous dopaminergic control of striatal synaptic plasticity. *Science,* **321**:848-851, 2008.

Simon, J. R., and Rudell, A. P. Auditory S-R compatibility: the effect of an irrelevant cue on information processing. *J Appl Psychol,* **51**:300-304, 1967.

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*, 2011, submitted.

Stürzl, W., Hoffmann, U., and Mallot, H.A. Vergence control and disparity estimation with energy neurons: Theory and implementation. *Lecture notes in computer science*, **2415**:1255–1260, 2002.

Sun, G. and Scassellati, B. A fast and efficient model for learning to reach. *International Journal of Humanoid Robotics*, **2**: 391-413, 2005.

Surmeier, D.J., Ding, J., Day, M., Wang, Z. and Shen, W., D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends inNeurosciences,* **30**:228-235, 2007.

Taylor, J., Olson, T. and Martin, W.N. Accurate vergence control in complex scenes. *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on.* 540 - 545, 1994.

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.

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.

Tsai, C.-C., and Brass, M. Does the human motor system simulate Pinocchio's actions? *Psychol Sci*, **18**:1058-1062, 2007.

Tsai, C.-C., Kuo, W.-J., Hung, D. L., and Tzeng, O. J. L. Action co- representation is tuned to other humans. *J Cognitive Neurosci*, **20**:2015-2024, 2008.

Tsang, E.K.C., Lam, S.Y.M., Meng, Y and Shi, B.E. Neuromorphic Implementation of Active Gaze and Vergence Control. *IEEE International Symposium on Circuits and Systems, ISCAS*, pages 1076 - 1079, 2008

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

van Dijck, H. and van der Heijden, F. Object recognition with stereo vision and geometric hashing. *Pattern Recogn. Lett*. **24**(1-3):137-146, 1999

Villgrattner, T. and Ulbrich, H., Design and Control of a Compact High-Dynamic Camera-Orientation System. *IEEE-ASME Trans. on Mechatronics*, **16**(2): 221-231, 2011.

Villgrattner, T. and Ulbrich, H., Optimization and Dynamic Simulation of a Parallel Three Degree-of-Freedom Camera Orientation System. *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2010,* 2829-2836, Taipei, Taiwan, 2010.

Vitay, J. and Hamker, F.H., A computational model of the inuence of basal ganglia on memory retrieval in rewarded visual memory tasks. *Frontiers in Computational Neuroscience,* **4**(13):1-18, 2010.

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

Volcic, R. and Lappe, M. Potentially purposeful actions divert overt attention. *Attention, Perception and Psychophysics*. Under revision. 2011

Voorn, P., Vanderschuren, L., Groenewegen, H.J., Robbins, T.W. and Pennartz, C., Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences* **27**:468-474, 2004.

Wallis, G. and Rolls, E. T. Invariant face and object recognition in the visual system. *Prog Neurobiol*, **51**(2):167-194, 1997.

Wang, X., Zhang, M., Cohen, I. and Goldberg, M. The proprioceptive representation of eye position in monkey primary somatosensory cortex. *Nature Neuroscience*. **10**: 640 - 646, 2007

Wang, Y. and Shi, B.E. Autonomous Development of Vergence Control Driven by Disparity Energy Neuron Populations. *Neural Computation*, **22**(3): 730-751. 2010.

Westheimer, G. and Mitchell, A.M.. Eye movement responses to convergence stimuli. *Archives of Ophthalmology*, **55**(6):848-856, 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.

Yang, Y. and Purves, D. A statistical explanation of visual space.*Nat. Neurosci*., **6**(6):632–640, 2003.

Zimmermann, E. Lappe, M. Eye position effects in oculomotor plasticity and visual localization. *J. Neurosci*. 2011 (in press).

Zimmermann, E., and Lappe, M. Motor signals in visual localization. *Journal of Vision*, **10**(6), 2. 2010.

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.*, 50:1328-1337, 2010

Zirnsak, M., Hamker, F.H. Attention alters feature space in motion perception. *J. Neurosci.*, 30(20):6882-6890, 2010