

COMENIUS UNIVERSITY IN BRATISLAVA  
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS  
DEPARTMENT OF APPLIED INFORMATICS



---

# LEARNING MULTISENSORY INTEGRATION AND COORDINATE TRANSFORMATION IN A SIMULATED HUMANOID ROBOT

*Master thesis*

---

2019

BC. MARTIN KELLNER

UNIVERZITA KOMENSKÉHO V BRATISLAVE  
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

UČENIE MULTISENZORICKEJ INTEGRÁCIE A  
TRANSFORMÁCIE SÚRADNÍC V SIMULÁTORE  
HUMANOIDNÉHO ROBOTA

*Diplomová práca*

Študijný program: Aplikovaná informatika  
Študijný odbor: Aplikovaná informatika  
Školiace pracovisko: Katedra aplikovanej informatiky  
Školiteľ: prof. Ing. Igor Farkaš, Dr.

Bratislava, 2019

Bc. Martin Kellner



Comenius University in Bratislava  
Faculty of Mathematics, Physics and Informatics

---

## THESIS ASSIGNMENT

**Name and Surname:** Bc. Martin Kellner  
**Study programme:** Applied Computer Science (Single degree study, master II. deg., full time form)  
**Field of Study:** Applied Informatics  
**Type of Thesis:** Diploma Thesis  
**Language of Thesis:** English  
**Secondary language:** Slovak

**Title:** Learning Multisensory Integration and Coordinate Transformation in a Simulated Humanoid Robot

**Annotation:** Sensory processing in the brain entails multisensory integration (combining cues into a single common underlying stimulus) and coordinate transformations (the change of reference frame for a stimulus, e.g. retinotopic to body-centered, effected through knowledge about an intervening variable, e.g. gaze position). This ability is crucial in cognitive robotics in order to endow robots with an ability to operate autonomously in the 3D space.

**Aim:**

1. Study the cognitive neuroscience literature related to sensory integration and the frames of reference (coordinate systems).
2. Implement and train an artificial neural network model on simulated data integrating visual and proprioceptive information, performing a coordinate transformation in hand-eye related task, using the iCub robotic simulator.
3. Evaluate and analyze the behavior of the trained model.

**Literature:** Makin J., Fellows M., & Sabes P. (2013). Learning multisensory integration and coordinate transformation via density estimation. PLOS: Comput. Biol., 9(4).  
Švec M., Farkaš I. (2014). Calculation of object position in various reference frames with a robotic simulator. In Proceedings of the 36th Annual Conference of the Cognitive Science Society, Quebec, Canada.  
Tikhanoff V., Fitzpatrick P., Nori F., Natale L., Metta G., & Cangelosi A. (2008). The iCub humanoid robot simulator. Advanced Robotics, 1(1), 22-26.

**Supervisor:** prof. Ing. Igor Farkaš, Dr.  
**Department:** FMFI.KAI - Department of Applied Informatics  
**Head of department:** prof. Ing. Igor Farkaš, Dr.

**Assigned:** 16.10.2017

**Approved:** 20.10.2017      prof. RNDr. Roman Ďurikovič, PhD.  
Guarantor of Study Programme

.....  
Student

.....  
Supervisor



Univerzita Komenského v Bratislave  
Fakulta matematiky, fyziky a informatiky

## ZADANIE ZÁVEREČNEJ PRÁCE

**Meno a priezvisko študenta:** Bc. Martin Kellner  
**Študijný program:** aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)  
**Študijný odbor:** aplikovaná informatika  
**Typ záverečnej práce:** diplomová  
**Jazyk záverečnej práce:** anglický  
**Sekundárny jazyk:** slovenský

**Názov:** Learning Multisensory Integration and Coordinate Transformation in a Simulated Humanoid Robot  
*Učenie multisenzorickej integrácie a transformácie súradníc v simulátore humanoidného robota*

**Anotácia:** Spracovanie senzorických podnetov v mozgu zahŕňa multisenzorickú integráciu (kombinovanie vodítok do jedného spoločného podnetu) a transformácie súradníc (referenčných rámcov, napr. z retinotopického na telo-centrický, ovplyvnenej modulačnou premennou, napr. pozíciou očí). Táto schopnosť je kľúčová v kognitívnej robotike, ak chceme vybaviť robota schopnosťou operovať autonómne v 3D priestore.

**Anotácia:** 1. Naštudujte si problematiku z kognitívnej neurovedy o multisenzorickej integrácii a referenčných rámcoch (súradnicových systémoch).  
2. Implementujte a natrénujte model umelej neurónovej siete, ktorá sa naučí integrovať vizuálnu a proprioceptívnu informáciu, vykonávajúc prepočet súradníc v úlohe týkajúcej sa ruky a očí, s využitím simulovaného robota iCub.  
3. Vyhodnoťte a analyzujte správanie natrénovaného modelu.

**Literatúra:** Makin J., Fellows M., & Sabes P. (2013). Learning multisensory integration and coordinate transformation via density estimation. PLOS: Comput. Biol., 9(4).  
Švec M., Farkaš I. (2014). Calculation of object position in various reference frames with a robotic simulator. In Proceedings of the 36th Annual Conference of the Cognitive Science Society, Quebec, Canada.  
Tikhanoff V., Fitzpatrick P., Nori F., Natale L., Metta G., & Cangelosi A. (2008). The iCub humanoid robot simulator. Advanced Robotics, 1(1), 22-26.

**Vedúci:** prof. Ing. Igor Farkaš, Dr.  
**Katedra:** FMFI.KAI - Katedra aplikovanej informatiky  
**Vedúci katedry:** prof. Ing. Igor Farkaš, Dr.  
**Dátum zadania:** 16.10.2017

**Dátum schválenia:** 20.10.2017      prof. RNDr. Roman Ďurikovič, PhD.  
garant študijného programu

.....  
študent

.....  
vedúci práce

**DECLARATION:** I hereby declare that this thesis is my own work and that all the sources I have used or quoted have been indicated and acknowledged as complete references.

.....

## Abstrakt

Slovenský abstrakt v rozsahu 100-500 slov, jeden odstavec. Abstrakt stručne sumarizuje výsledky práce. Mal by byť pochopiteľný pre bežného informatika. Nemal by teda využívať skratky, termíny alebo označenie zavedené v práci, okrem tých, ktoré sú všeobecne známe.

**Kľúčové slová:** jedno, druhé, tretie (prípadne štvrté, piate)

## **Abstract**

Abstract in the English language (translation of the abstract in the Slovak language).

**Keywords:**

# Contents

<b>1</b>	<b>Introduction</b>	<b>9</b>
<b>2</b>	<b>Cognitive focus</b>	<b>10</b>
2.1	Reference frames . . . . .	10
2.2	Sensorimotor transformations . . . . .	12
2.2.1	Spatial Transformations for Eye–Hand Coordination . . . . .	13
2.3	Gain fields . . . . .	20
2.4	Use of artificial neural network . . . . .	24



# List of Figures

1	Schematic illustration of the spatial representations of objects in frames of reference. . . . .	11
2	Gaze-centered encoding of reach space. . . . .	15
3	Gaze-centered pointing performance in humans for targets in near and far space. . . . .	16
4	Conceptual scheme for spatial transformations in eye– hand coordination. . . . .	19
5	Working principle of gain fields, based on Zipser & Andersen (1988). .	21
6	A coordinate transformation performed by the visual system. . . . .	22
7	Visual responses that are gain-modulated by gaze angle. . . . .	23

# List of Tables

# List of Abbreviations

RF/RFs      Reference frames

PMd          Porsal premotor cortex

RCF/RCFs    Receptive fields

# Introduction

# Cognitive focus

## 2.1 Reference frames

In order to gain a better understanding of the term *reference frame* or *frame of reference*, consider the following situation. On a moving train there is a passenger who lets a book fall down and another person watching the situation is standing on the ground. From the perspective of the passenger the book is falling straight down, but the observer (the man standing on the ground) sees the book dropping down along a curved path because of the movement of the train. This situation has shown that the description of a physical phenomenon depends on the position from where the case is observed. Using reference frames (RF) on the previously illustrated example, we can say that the book is falling straight down to the train's RF but it is dropping along a curved path in the earth's RF. It can be said about the state of motion of the passenger that it is stationary in the train's RF but moving in the earth's RF (see the figure 1) (Soechting & Flanders (1992)).

RFs are very often used by physicists and engineers but this term has also been adopted by neuroscientists. In the mathematical sense, the RF is similar to a coordinate system that is defined by a set of axes and an origin. The origin can be anywhere in space and the orientation of axes can be chosen arbitrarily because RF is only characterized by the state of motion relative to an object. Therefore, the train's RF that has been mentioned in the illustrated example has the same state of motion as the train, and any point in the RF can be defined by its position along each of the coordinate axes. Alternatively, the location of a point can be defined by way of vectors, where a vector has two properties: magnitude and direction; the magnitude is the length of a line segment between the origin and the point, and its

direction is from the origin to the point (Soechting & Flanders (1992), Kumar & Barve (2002)).

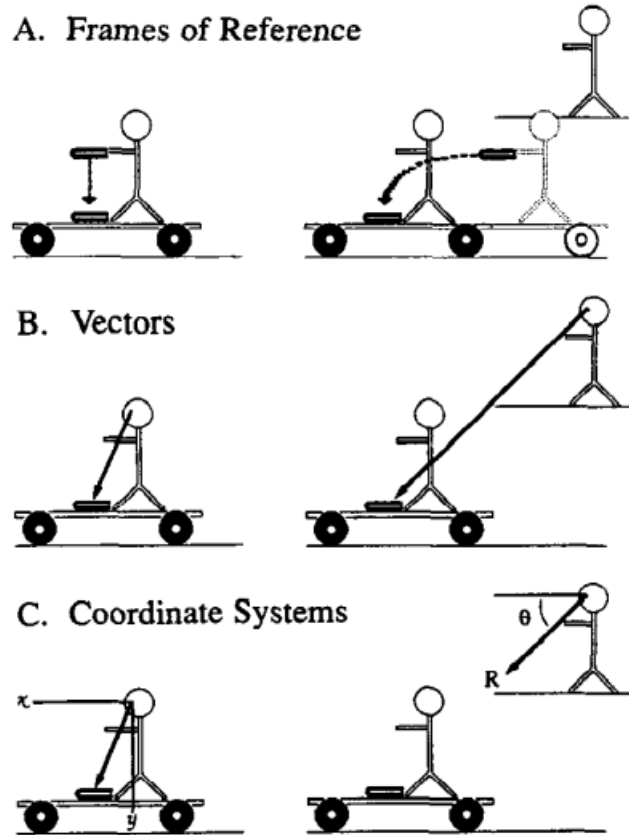


Figure 1: Schematic illustration of the spatial representations of objects in frames of reference (A), vectorially (B) and by coordinate systems (C). On the left, the RF moves with the passenger; on the right, the observer’s RF is fixed to the earth. With the description adopted from Soechting & Flanders (1992)

In order to switch the RF into a term in neuroscience, we have to make some changes against the mathematical definition because neural systems do not report to the position of an object as a vector or as coordinates. Instead, neurons encoding visual space each reports to a restricted area of space known as *receptive field* or *response field* (RCF), thus, the firing activity of the neurons is changed depending on stimuli coming from their RCFs (Batista (2002)). If we want to identify in which RF, for example, a neuron encodes spatial information about the location of an object, then the firing activity of the neurons should stay the same as long as the image of the object falls on the same locus on the retina and stays constant. Once that happened we can say the location is encoded in a *retinocentric* RF (Soechting & Flanders (1992)).

Different brain regions encode spatial locations in different RFs. In neuroscience RFs are mostly divided into two main groups: *ego-centric* and *allocentric* RFs. In an egocentric RF, the location of the object refers to the observer. At a neural level responses of the neural population are attached to a reference point (Committeri et al. (2004)). For instance, an eye-centred RF moves together with the eyes. Spatial locations are mostly encoded in egocentric RFs, especially space coding neurons in the parieto-frontal cortex are associated with RFs centred to the eye, the head or the hand (Colby (1998)). They have found neurons that encode space in RFs centred to parts of the body, in the monkey's posterior parietal cortex and in connected regions of the premotor cortex (Cohen & Andersen (2002); Colby (1998)). Information encoded in an allocentric RF is described with respect to other objects, for example, the book is on the table. Coding of space in allocentric RFs is not so well-studied as egocentric RFs, but it has shown (Marshall et al. (2000); Galati et al. (2000)) that in the posterior parietal cortex, in the dorsal premotor cortex (PMd) and in early visual areas neurons report during object-based spatial judgement. Also it has been proven that during the solving of a complex task, such as landmark knowledge, orientation in large-scale space and navigation, some operations refer to both types of RFs, allocentric and egocentric, which are difficult to untangle (Committeri et al. (2004)).

## 2.2 Sensorimotor transformations

If humans, animals or insects want to carry out basic operations to interact with the world they use their bodies and senses. To manipulate with an object, moving in space or doing other daily activities, the brain must continuously update the internal representation of the world, deal with restrictions that may occur while performing an operation and change body configuration to achieve the goal of the action. For instance, during the reaching of an object the human brain must take into account the location of the target and the current position of the particular hand to generate the right movement for the hand. One non-neuroscientist might say that the operations as above mentioned are not very complicated because we do not need to put a lot of effort into getting the goal but when we look closely at the

neural level the task is more complicated as it would seem. The information about the location of an object can be produced by stimuli from different modalities and each of these modalities provides neural information that is encoded in a different RF. Therefore, neuroscientists focus their research to clarify the following questions: How is the information represented in a RF at the neural level? How is the information from different modalities combined to encode a correct representation of space? How is the information in a particular RF transformed to another one? Coordinate transformation in neuroscience is a term used for the transformation from one RF to another. In the next section we closely discuss the eye-hand coordinate transformation that is crucial for many human daily activities.

### **2.2.1 Spatial Transformations for Eye–Hand Coordination**

The eye-hand coordination is reviewed by Crawford et al. (2004) and Blohm et al. (2009) in very detail. The authors have brought assumptions about using gaze-centred representations during the process of coordinate transformation that is needed to perform such tasks as reaching or grabbing an object. In this chapter we closely discuss their conclusions.

In order to reach an object, the brain needs information about the location of the object being reached and the position of this object has to be encoded in relation to the body part performing the operation. Arms of primates are anchored to shoulders and the spatial information about an object comes primarily from projections on retinas, therefore, the brain must perform a coordinate transformation between at least two representations. First of them is the representation specifying the object's location relative to retinas and the second representation within which the activations of arm muscles must be determined to reach the target. In this section we discuss how the brain deals with the transformation of visual stimuli into hand motion commands and what processes are included as grabbing, reaching or manipulation with objects.

A robot system controlling its arm to reach or to grab an object is usually simpler than the human control system. Such robotic system is ordinarily reduced only by using visual feedback. Basically, driving the hand to a point that is seen by the visual system is carried out only by comparing the current position of the hand with the



target position. This solution may be sufficient because the speed of visual feedback of such robotic system is not so dramatically limited as the brain. The limitation of the robotic visual feedback is caused only by processing the time of the processor and the speed of the electrical flow, whilst the speed of neural conduction and processing time in a real primate brain is not adequate to perform a fast hand movement. Thus using just visual feedback would probably cause that the hand would be out of the field of vision before a new visual feedback would arrive to accurately update the movement. Consequently, the brain of primates cannot use only visual feedback driving the hand's movements and it must find another way to carry out correct and fast movements. The answer is the usage of internal models of the physical system representing the world based on initial conditions and eye-hand coordinate transformation performing in the feedforward way. Nevertheless, we cannot say that visual feedback does not contribute to the guiding hand movements because visual feedback actually helps to achieve the best performance of grabbing a goal and is also essential for dealing with unexpected events or danger that might occur during the performance of a particular movement.

We still do not know all the processes that are included in the brain to produce a correct transformation from eyes to hands and how these processes are exactly carried out at the neural level. Regardless, there are several conclusions made by neuroscientists that we discuss in the following two conceptual steps.

### **Early visual representation for arm movements**

For operations as coordinate transformation we usually describe internal models in the brain by using the concept of RF (2.1). The final RF for eye-arm coordination is associated with points in the upper arm and shoulder and not with a hand-centred representation as we could mistakenly think. The first stimuli are controlled in a retinal RF and the first problem occurs here: *gaze shifting*. Every time when the eyes (or/and head) shift(s) with the gaze, the visual relationship between the sensory apparatus and the external world changes (Hallett and Lightstone 1976). One solution, how the brain could deal with this problem, is to wait until the gaze shifting finishes and then update its visual information (O'Regan and Noe 2001) but that might cause that the original target of interest would be moved into a less-

sensitive peripheral retina or even out of the visual field. It is good to realise that this solution would produce also long lags in processing the time and redundant visual computations. Therefore, the necessary representations must be stored for future action either in an eye-independent form, or in such form that is internally updated if the gaze is shifted, regardless of the fact that they were caused by movements of eyes or the head (Duhamel et al. 1992).

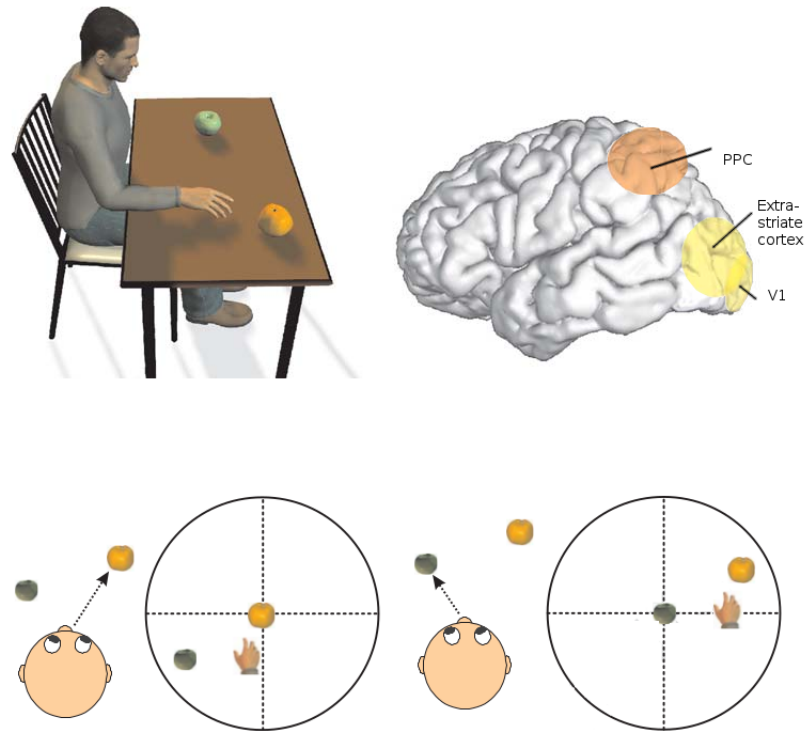


Figure 2: Gaze-centered encoding of reach space. (a) Drawing depicting the egocentric visual directions of the hand, an orange, and an apple, as shown by the gray arrows. (b) Side view of a human brain showing areas (highlighted in yellow and orange) that encode reach space in gaze-centered coordinates: V1, striate cortex; PPC, posterior parietal cortex. (c) The consequence of an eye movement on the gaze-centered representation of the visual field. The head/eye diagrams depict current gaze position; the circles represent the gaze-centered representation of this visual scene (dotted lines represent visual horizontal and vertical axes and intersect at the fovea). If the person in the upper diagram looks at the orange, the hand and apple are represented in the left visual field (upper circle). In contrast, if the person fixates the apple (lower diagram), the orange and hand are now represented in the right visual field (lower circle). If the orange and the hand were no longer visible when the eye movement occurred, the brain would need to remap their position by taking the intervening eye movement into account. With the description adopted from Blohm et al. (2009)

The next raising question is: How the system in the human brain, which appears to be used for early planning of pointing or reaching, stores early motor representa-

tions during eye movements? In order to answer the question, Henriques et al., 1998 performed the following experiment. Human subjects that were carrying out the experiment, were asked to point toward a location using one hand in total darkness. At the beginning of the experiment, the target location that was pointed to, was centred to subject's fovea. Then subjects shifted the gaze by moving their eyes and then pointed toward the remembered location.

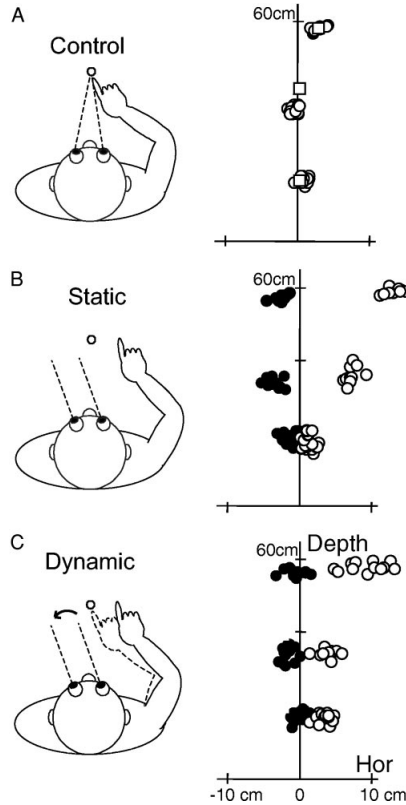


Figure 3: Gaze-centered pointing performance in humans for targets in near and far space. Left column: 3 tasks, where subjects either (A) look directly toward the target before pointing (control task) or (B) view the target peripherally before pointing (Static Condition) or (C) foveate the target, then shift their gaze and then pointing (Dynamic Condition). Right Column: final fingertip positions (circles) in the horizontal plane of one subject these conditions. Squares represent the actual target locations of the two reaching targets and the fingertip location for pointing toward the continuously illuminated pointing target. In static and dynamic tasks, open circles indicate 20 leftward eye fixation; solid circles represent data for 20 rightward eye fixation. Targets were located at 2 m, 42 cm, and 15 cm. With the description adopted from Crawford et al. (2004)

Authors of the experiment tried to find out whether the responses of such a pointing task are affected by an intervening eye displacement. They compared them to the pointing to the remembered foveal targets or pointing to the retinally peripheral targets. If subjects were pointing toward the location using an updated gaze-centred representation, pointing behaviour would echo pointing to peripheral targets. Whereas pointing by using a nonretinal representation would cause no effect. As shown using the figure 3, this experiment has cleared up the question about the early motor representation and it has supported the idea of using an updated gaze-centred RF. Other studies have also come up with findings where during pointing to auditory and proprioceptive targets the gaze-centred updating has been also

recorded (Pouget et al. 2002). Evidence of coding the spatial information of an object in a gaze-centred RF can be also found in both the early visual pathway (i.e., retina, lateral geniculate nucleus, striate cortex) and also later in the visual pathway (e.g., extrastriate and parietal cortex).

The gaze-centred representation naturally moves depending on the gaze because the retinal projection of the project depends on the gaze. Therefore, the positions of the target being reached can be defined by the target's direction (horizontal and vertical angular eccentricity) relative to *fovae* and by the distance between the goal and the eyes. Early visual areas carry out the computation of target direction as seen by a virtual eye which we can imagine to be placed between the right and the left eye, whereas the required distance is determined by taking into account monocular information (accommodation, relative object size, shading, perspective, etc.) and binocular information (retinal disparity and convergence). Later, during the early movement planning, the signals that contain this information seem to be merged into a single gaze-centred representation of the space in the posterior parietal cortex.

As we have mentioned above, the gaze-centred representation must be updated to maintain a stable representation of the world. Therefore, the brain must update this representation all the time the eyes are moving. This process, which performs during the rotation of eyes, is known as *updating* or *remapping* and it also occurs during the movements of the head and the body. An illustrating example is in the figure1(c). The observer originally looks at the orange, and the apple is placed left-down relatively to the orange, thus the orange corresponds to the centre of the gaze-centred representation, and the apple is located on the lower-left quadrant of the visual field. A fast orientation movement approaching the apple causes that the objects are remapped by the same rotation as the movement but in the opposite direction. That means the apple is mapped at the centre and the orange is located on the upper-right quadrant. It has also been shown that the updating can be achieved even if the vision is removed, thus only by using information about the eye movements.

## Developing the reach plan

In order to correctly generate the reachable plan, the information about the initial position of the hand must be included. Buneo et al, 2002 have introduced a suggestion that the comparison between initial hand location and gaze-centred representation of the visual target is done earlier and in a gaze-centred RF, and even if the hand is out of the visual field. That implies that the proprioceptive signals carrying the information about hand location must be transformed into gaze-centred coordinates. These kinds of signals were found in the process of a gaze-centred transformation in parietal area 5, but these findings regarding the earlier comparison do not claim that the next RF transformation is not required, or needed to reach the target.

The next important problem, the brain has to deal with in order to accurately generate and perform a reaching plan, is how the ego-centre is translocated during the head rotation. This problem has to be taken into account because the rotation of the head causes eye's translation with respect to the shoulder. Ignorance of eye translational aspect would cause erroneous reach pattern at noncentral head position.

A summary of conceptual and physiological models of visually guided reaching movement is described in the figure 4. The first stage of the figure corresponds to the fact that the 3D representations of target direction are stored and maintained in a RF. The second stage shows an illustration of the transformation according to the above-mentioned Buneo et al. (2002) schema to compute the hand displacement in a RF coordinate. There are other RF transformations from gaze coordinates to shoulder representations that are not included in the figure but they are necessary to reflect the eye-head-shoulder system. Nevertheless, such a model may be useful to tell us which signals are required and used for such a transformation but it is obvious that all such kinds of explicit intermediate representations are not used by the brain directly, and they also do not tell us how and where the mentioned signals are coded.

To design a complete 3D reachable plan, the brain must know the amplitude of the desired movement and the direction. It has shown that in the dorsal premotor cortex the direction and the amplitude are encoded together inside the same neurons,

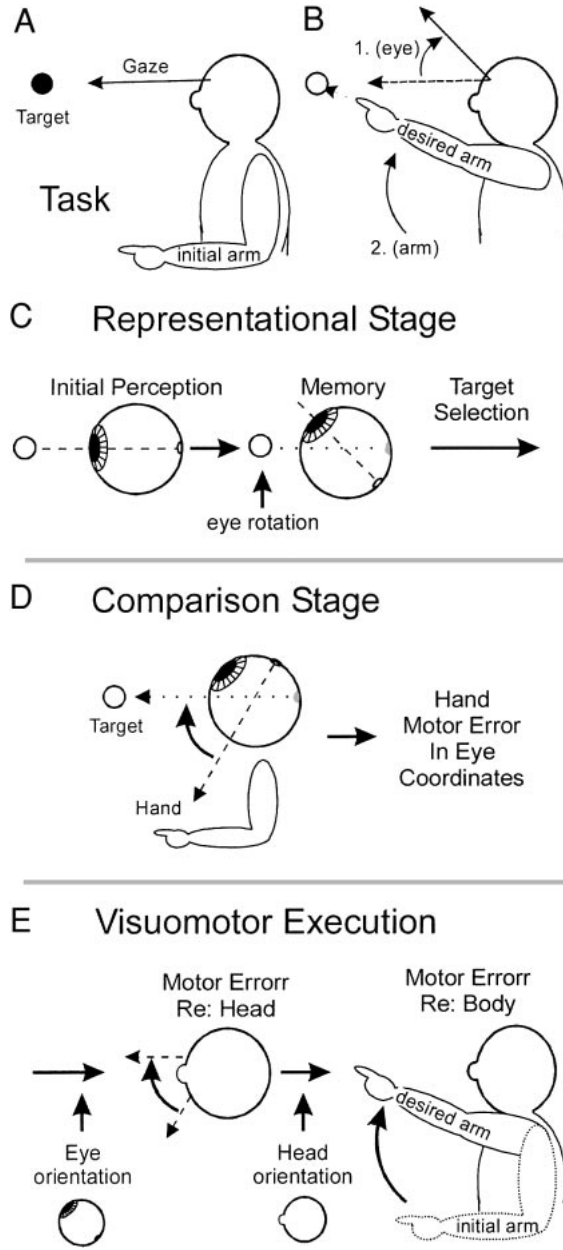


Figure 4: Conceptual scheme for spatial transformations in eye–hand coordination. To illustrate the model, consider the following “task”: a subject looks at a briefly flashed target ( F ) with the arm at resting position ( A ). Then ( B ) the subject makes 1) an upward eye movement, followed by 2) a reaching or pointing movement toward the remembered target location ( E ). We hypothesize that the brain uses the following stages to do this. C: an early representational stage. Target location is stored in eye coordinates, such that this representation ( E ) must be counterrotated (updated) when the eye rotates. D: comparison stage. Updated target representation ( E ) is compared with an eye-centered representation of current hand location to generate “hand motor error” in eye coordinates (Buneo et al. 2002). E: visuomotor execution stage. “Hand motor error” signal is rotated by eye orientation and head orientation (or perhaps by gaze orientation) to put it into a body coordinate system appropriate for calculating the detailed inverse kinematics and dynamics of the movement. With the description adopted from Crawford et al. (2004). This last stage would also have to include internal models of the geometry.

but there is also an evidence that the direction and the amplitude might also be encoded independently. One explanation of this contradiction might be that, at the neural level, amplitude and direction are encoded together but are used by different mechanisms to generate muscle activation. In must be taken to account the rotation of the head and body and also values of correct forces for specific muscles. Amplitude and direction are properties of vectors, therefore, we can reformulate the problem of designing a reachable plan into the terminology of vectors as that is often done.

The conclusion of the problem of developing a reachable plan is that the brain must design a hand movement vector, which depends on the difference between

the hand location and the location of the target. The areas in the brain that are responsible for calculating the vector, have to contain encoded spatial information of the hand and the position of the goal. In addition, the spatial information must be represented in the same RF. To determine the movement vector, the particular areas might use information from different sources, and also perform multisensory integration in order to get the most likely estimation of both the hand and the target location. One of the approaches that help us to examine how the population performs the task of coordinate transformation is using artificial neural networks. During studying of neural encoding of a cognitive function, we very often focus on changing response magnitude of neurons known as *gain field* what we are discussing in the next section.

## 2.3 Gain fields

The term *gain field* comes from Andersen & Mountcastle (1983). The authors tested the visual receptive field for a specific neuron at different eye positions. The receptive field of a neuron or a neural population can be described as a specific region in sensory space of which stimuli can produce an influence on the activity of the single neuron or on the neural population. For instance, activations of neurons encoding the gaze-centered representation of a target's location should be constant while gaze angles stay constant, contrariwise, firing activities of the neurons should change with every motion of the eyes. During the mentioned experiment the frequency of the neural action potential was changing in a manner of multiplication the frequency by a gaze angle that is scaled by a constant. Both of the shape and the locations of the visual receptive field were unchanged, only the receptive field was scaled by some *gain* factor. This phenomenon was the first time characterized in neurons inside LIP and visual area 7A. Another study (Zipser & Andersen (1988)) examined the task of coordinate transformation from visual target position on retinas and signals of gaze position into the position represented by the target location in a space-fixed RF by an artificial neural network ( ANN ). The detailed analysis of their neural model showed that the network was able to develop visual receptive fields modulated by the position of the eyes in a very close way compared to the modulation in parietal

cortex. The figure 5 illustrates the results of their study.

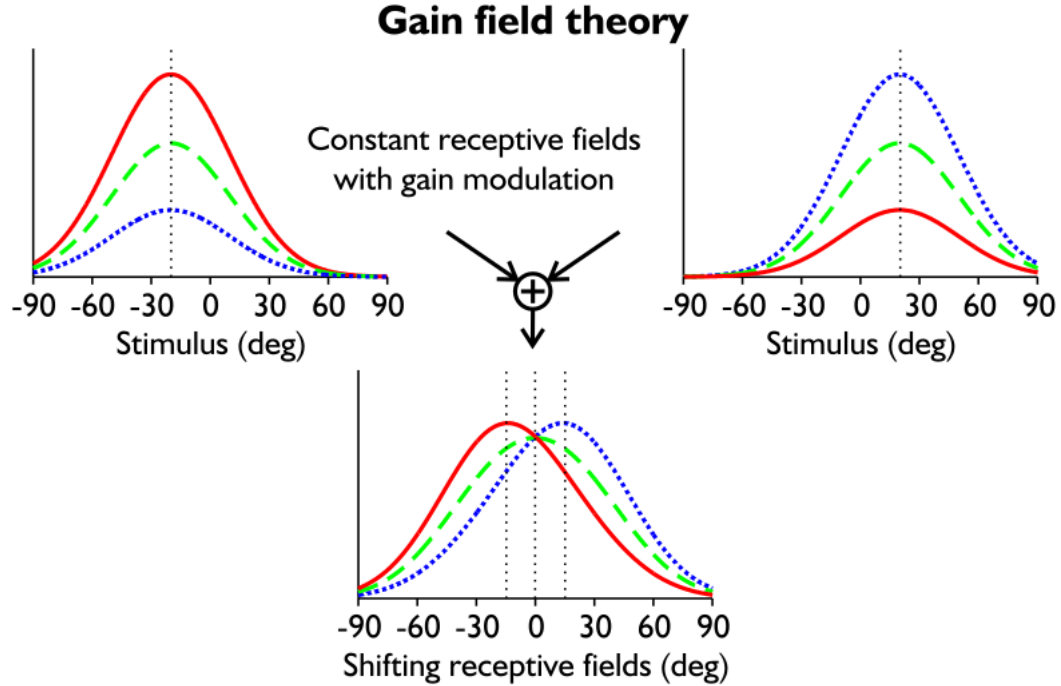


Figure 5: Working principle of gain fields, based on Zipser & Andersen (1988). The upper part of the panel shows the hypothetical receptive fields of two neurons that are gain modulated (e.g., by eye or hand position) in opposite ways without shifting. For example, the three lines in each graph could represent visual receptive fields mapped relative to gaze at a leftward eye position (red solid line), a central eye position (green dashed line), and a rightward eye position (blue dotted line). Here, eye position modulates the strength of response of two neurons, but does not cause them to shift. However, summation of these two gain-modulated neural responses results in shifting receptive fields in the output, e.g., eye position (or in other cases hand position) has shifted the receptive field. With the description adopted from Blohm & Crawford (2009).

At the neural population level, gain fields modulate the responses of every particular neuron in a whole population responding to the same receptive fields, and therefore, the total output of the population can be increased or decreased that allows to regulate the total strength of the receptive field. In the next steps, the brain can integrate such an output in various ways and use them in next computations. For instance, something similar to the output of the above mentioned ANN could be hypothetically compared with information about the current hand's location to produce the hand movement to reach a target. It has been observed that the effect of gain modulation is also produced by other types of signals, and gain fields were found in many other brain areas (Blohm & Crawford (2009)).

In order to demonstrate the role of gain modulation with the performance of the



task of coordinate transformation in the brain, we will discuss again the results that have been introduced by Zipser & Andersen (1988). As mentioned, they trained a neural network to solve coordinate transformation from the input signals, gaze direction and retinal location to output stimuli in a body-centred RF. The model was trained using one of the major-used training algorithms in machine learning known as *backpropagation* (backprop). This algorithm allows learning of properties of *hidden* units (neurons) in order to map inputs to corresponding outputs with the effort to achieve as low error as possible. All of the output and input representations of stimuli were composed to match the measured response of neurons in the brain, thus the model was able to perform a coordinate transformation in the similar way as it is made by the visual system (see the figure 6). After a successful training, the authors have shown that hidden units developed gaze gain fields similar to those in PPC. This experiment has brought a significant contribution that supports the role of gain modulation within the process of solving a coordinate transformation in the brain (Salinas & Sejnowski (2001)).

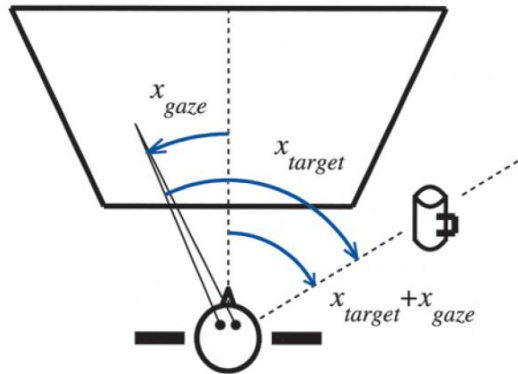


Figure 6: A coordinate transformation performed by the visual system. While reading a newspaper, you want to reach for the mug without shifting your gaze. The location of the mug relative to the body is given by the angle between the two dashed lines. For simplicity, assume that initially the hand is close to the body, at the origin of the coordinate system. The reaching movement should be generated in the direction of the mug regardless of where one is looking, that is, regardless of the gaze angle  $x_{gaze}$ . The location of the target in retinal coordinates (i.e., relative to the fixation point) is  $x_{target}$ , but this varies with gaze. However, the location relative to the body is given by  $x_{target} + x_{gaze}$ , which does not vary with gaze. Through this addition, a change from retinal, or eye-centered, to body-centered coordinates is performed. With the description adopted from Salinas & Sejnowski (2001).

Other researches also took an effort to study similarities between real and artificial neurons, especially between their responses. Xing & Andersen (2000) trained a more complex neural model to represent two consecutive saccades separated by a de-

lay interval. They used a recurrent neural network that is able to learn a sequential task through their internal *memory*. Input data combined a visual map in retinal coordinates, an auditory map in head-centred coordinates and eye-position units. The model was taught to produce outputs that encode simultaneous movements of both eyes (Xing and Andersen, 2000). After the training, they demonstrated that their network developed gain fields and sensory and memory responses matched those of LIP neurons (Salinas & Sejnowski (2001)).

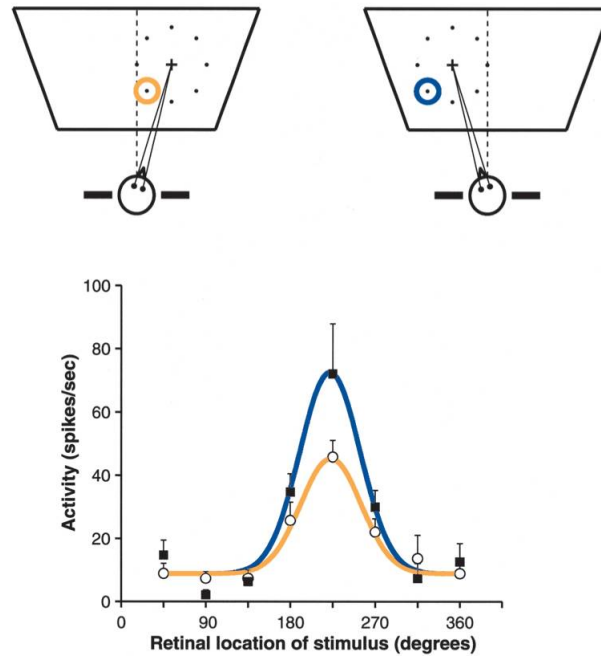


Figure 7: Visual responses that are gain-modulated by gaze angle. The response of a parietal neuron as a function of stimulus location was measured in two conditions, with the head turned to the right or to the left, as indicated in the upper diagrams. In these diagrams, the cross corresponds to the location where gaze was directed, called the fixation point; the eight dots indicate the locations where a visual stimulus was presented, one location at a time; and the colored circles show the position of the recorded neuron’s receptive field. This was centered down and to the left of the fixation point. In the diagrams, the rightmost stimulus corresponds to 0 degrees, the topmost one to 90 degrees, and so forth. The dashed line indicates the direction straight ahead. The graph below plots the neural responses in the two conditions, indicated by the corresponding colors. The continuous lines are Gaussian fits to the data points. When the head is turned, the response function changes its amplitude, or gain, but not its preferred location or its shape. With the description adopted from Salinas & Sejnowski (2001).

Salinas and Abbott (1995) studied the relationship between the coordinate transformation described in the figure 6 and gain modulation. Their research was focused on neurons that are gain-modulated by the gaze and drive downstream neurons in another neural population included in the process of generating arm movements.

They considered a neural population that responds to the visual stimulus at the retinal location  $x_{target}$  and is modulated by the gaze angle  $x_{gaze}$  that constitutes gain modulation. Then all responses  $r$  of such a population can be described by the following equation:

$$r = f(x_{target} - a) g(x_{gaze}), \quad (2.1)$$

where  $g(x_{gaze})$  is the function of which result is the gain field of particular neuron, and  $f(x_{target} - a)$  represents the response of neurons as a curve with a single pick (see the figure 7). Then, downstream neurons driven by a population as mentioned must have responses  $R$  theoretically formed by the following function of  $x_{target}$  and  $x_{gaze}$ :

$$R = F(c_1 x_{target} + c_2 x_{gaze}), \quad (2.2)$$

where both constants  $c_1$  and  $c_2$  depend on the synaptic weight, and the receptive field of a downstream neuron is represented as the peaked function  $R$ .

This mathematical formulation is only a simplification of the problem but the authors found conditions under which this can happen and they experimentally confirmed that the downstream neurons encode the sum of  $x_{target} + x_{gaze}$ , which means the downstream neurons that must encode target in a body-centred RF respond as a function of  $x_{target}$  that shifts if  $x_{gaze}$  changes (Salinas and Abbott (1995)). A similar shift has been demonstrated in several next cortical areas associated with a variety of coordinate transformations (Salinas & Sejnowski (2001)).

## 2.4 Use of artificial neural network

# Bibliography

- Andersen, R. & Mountcastle, V. (1983), ‘The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex’, *The Journal of Neuroscience* **3**(3), 532–548.
- Batista, A. (2002), ‘Inner space: Reference frames’, *Current Biology* **12**(11), 114–117.
- Blohm, G. & Crawford, J. D. (2009), ‘Fields of gain in the brain’, *Neuron* **64**(5), 598–600.
- Blohm, G., Khan, A. & Crawford, J. (2009), ‘Spatial transformations for eye–hand coordination’, *Encyclopedia of Neuroscience* p. 203–211.
- Cohen, Y. E. & Andersen, R. A. (2002), ‘A common reference frame for movement plans in the posterior parietal cortex’, *Nature Reviews Neuroscience* **3**(7p), 553–562.
- Colby, C. (1998), ‘Action-oriented spatial reference frames in cortex’, *Neuron* **20**(1p), 15–24.
- Committeri, G., Galati, G., Paradis, A., Pizzamiglio, L., Berthoz, A. & Lebihan, D. (2004), ‘Reference frames for spatial cognition: Different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location.’, *Journal of cognitive neuroscience* **16**, 1517–1535.
- Crawford, J., Medendorp, P. & Marotta, J. (2004), ‘Spatial transformations for eye-hand coordination’, *Journal of neurophysiology* **92**, 10–9.
- Galati, G. and, L. E., Vallar, G., B. A. P. L. & Bihan, D. (2000), ‘The neural basis of egocentric and allocentric coding of space in humans: a functional magnetic resonance study’, *Experimental Brain Research* **133**(2p), 156–164.
- Kumar, A. & Barve, S. (2002), *How and why in basic mechanics*, Orient BlackSwan Universities Press.
- Marshall, J., Fink, G., Shan, N., Weiss, P., Halligan, P., Grosse-Ruyken, M., Ziemons, K., Zilles, K. & H., F. (2000), ‘Line bisection judgments implicate right parietal cortex and cerebellum as assessed by fmri’, *Neurology* **54**(6), 1324–1331.

- Salinas, E. & Sejnowski, T. J. (2001), ‘Book review: Gain modulation in the central nervous system: Where behavior, neurophysiology, and computation meet’, *The Neuroscientist* **7**(5), 430–440.
- Soechting, J. & Flanders, M. (1992), ‘Moving in three-dimensional space: Frames of reference, vectors, and coordinate systems’, *Annual Review of Neuroscience* **15**(10), 167–191.
- Xing, J. & Andersen, R. A. (2000), ‘Memory activity of lip neurons for sequential eye movements simulated with neural networks’, *Journal of Neurophysiology* **84**(2), 651–665.
- Zipser, D. & Andersen, R. A. (1988), ‘A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons’, *Nature* **331**(6158), 679–684.