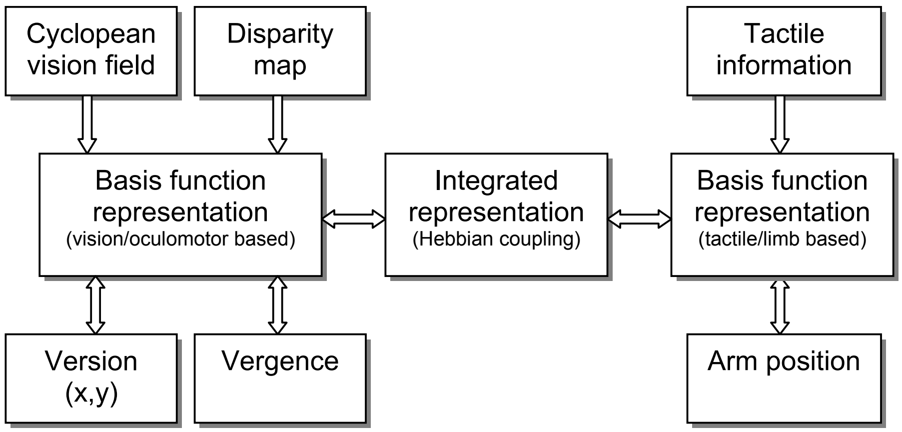
Model of visuomotor spatial awareness – WP4 – September 09 meeting update



The combination of eye and arm control into a common representation of spatial awareness relies on the fact that eye and arm movements usually go together, i.e., we fixate an object before, or while, we reach towards it. Such combinations of looking and reaching for the same target can be used to establish a common representation performing coordinated movements aimed at bringing both representations in register.

In the proposed framework (image above), eyes and arms are treated as separate effectors that receive motor control via different specific representations which, however, combine to form a unique shared representation of visuo-motor awareness. Each of the two representations is maintained through a radial basis function network, chosen because of its biological plausibility and its properties especially suitable to reference frame transformations.

The left side of the schema concerns the integration of stereoptic visual information with oculomotor control. Among the possible alternatives for representing binocular information we favor the composition of a cyclopean image representation with a disparity map (under the assumption that the correspondence problem is already solved), over the option of having separate left and right retinotopic maps. Similarly, considering that we are modeling extrastriate and associative visual areas, it is plausible to assume that gazing direction is represented by version and vergence angles instead of the two explicit eye positions. This scheme allows us to transform ocular movements and stereoptic visual information to a body-centered reference frame but also, when needed, elicit the eye movements that are necessary to foveate on a given visual target. Logpolar image representations can be used to simulate foveal magnification.

The second basis function representation links the movements of the arm with another body-centered representation of the space, based on the tactile exploration of the environment instead than on vision. It remains an open issue to determine how arm position is to be coded. At least three degrees of freedom are required, and six would allow to code also for reaching direction and not only for target position. In any case, the basis function representation is to be used for coding the position of objects in the peripersonal space considered as potential targets for reaching movements, and the vectors corresponding to the actual movements can be extracted similarly to what is done for ocular movements.

The first implementation stage consists in establishing the structure of the basis function representations, to make them suitable for the whole model. Then, Hebbian learning applied to actual experiments on the robotic setup will be employed to put the two representations in register with each other in order to obtain a third one, constituted by links between the two basis function maps. The integrated representation that allows to contextually represent the peripersonal space through different vision and motor parameters is thus never made explicit, but rather emerges from the two separate representations thanks to the interaction of the agent with the environment. Coordinated reach/gaze actions will be the tool used to integrate and match the two maps. This learning process is the normal behavior of the agent, and constitutes the most fundamental component of its basic capability of interacting with the world and contextually updating its representation of the world itself.

There are three fundamental aspects that constrain the development of the above schema. First of all, the modeling has to be done keeping in mind the requirements of the robotic implementation and the actual experimental part. In fact, overly complex and long learning frameworks have to be avoided such as for example reinforcement learning processes, which can be hardly applied on real robots. On the other hand, it is easy to fall into trivial problems. For example, for combining two different 3D reference frames, the exact matching of three points in space, and thus a sequence of only three reaching/gazing movements can be enough. Learning should instead be progressive and movement errors should be used to correctly update the global representation. Ideally, this stage could be used also to improve thesingle basis function representations, thus obtaining a sort of ``self-supervised learning'' framework, in which the different modalities supervise each other, and eye and arm movements both improve, and obtain together a precise visuomotor representation of the surrounding space. In any case, the actual learning procedure has thus to be clearly defined during the first stages of the implementation, not later on.

The second point is the use of data and insights from WP5, regarding V6A neurons in macaques and saccadic adaptation in humans (see D4.2a). Of course, the model can’t include or respect all aspects of the neuroscience experiments, but will have to clearly take into account/reproduce some fundamental ones. This second issue is less critical than the first one for the beginning of the modeling process, but still should not be disregarded or excessively postponed.

Third, the whole framework has to be suited to the final experimental setup which includes a human-robot interaction setup inspired on the experiments by WWU. The proposed increasingly complex capabilities of the robot according to such a framework are:

- Human gazes at one of many objects, robot reacts reaching/grasping it; the start signal can be given by time of gazing, or by gaze field discrimination. In all stages of the experiment it is fundamental to define a communication code to make the robot understand different cues given by the human subject.

– Human gazes subsequently at different objects/positions in space, that the robot has to grasp/reach; the robot is always looking at human gaze, in order to predict position/identity of next goals, and thus has to grasp/point an object without looking at it.

– More advanced tasks could be: 1) Human gazes on an object and might reach for it or not. The robot performs the reach to the gazed object, but if the human starts moving his arm the robot has to abort its movement. 2) The robot is instructed to move away the hand from an object if the human is approaching it or fixating it.